Volume 27 Number 1 March 1976
BJPIA5 27 (No. 1) 1-104 (1976) ISSN 0007-0882

The British Journal for the Philosophy of Science

18

Published for the
British Society for the Philosophy of Science
by Aberdeen University Press

THE BRITISH JOURNAL FOR THE PHILOSOPHY OF SCIENCE

Editors: JOHN WATKINS and JOHN WORRALL

Assistant Editors: COLIN HOWSON and ELIE ZAHAR

Subscriptions

The *Journal* is published quarterly. The subscription price is £7.00 net (U.S.A. and Canada U.S. $22.00) per volume post free; single issues £2.50 net (U.S.A. and Canada U.S. $8.00) plus postage. The subscription is £4.50 (U.S.A. and Canada U.S. $14.00) for members of the British Society for the Philosophy of Science, who should order direct from the publisher. (Membership of the Society is international and is by no means confined to people resident in the United Kingdom.) Orders may be sent to any bookseller or direct to Aberdeen University Press, Farmers Hall, Aberdeen, Scotland AB9 2XT.

Notes to Contributors

Articles, correspondence and books for review should be sent to John Worrall at the Department of Philosophy, London School of Economics, Houghton Street, Aldwych, London WC2A 2AE. Articles should be accompanied by abstracts of about 100 words.

Material is accepted for publication in the *Journal* on the condition that copyright is assigned to the British Society for the Philosophy of Science.

Three copies of all contributions should be sent, typed on one side only with wide margins and double-spaced. Footnotes, which should be as few as possible, should be typed separately, double-spaced at the end of the text. They should be referred to by consecutive numbering throughout the text, in which they will be inserted in page proof. Bibliographical references should be listed alphabetically at the end. The style of references should follow current *Journal* practice.

Long quotations are printed in small type without quotes, and should be so marked in the typescript. Single quotes should be used to form names of words and for shorter quotations; double quotes for other purposes. Diagrams, tables, and illustrations should be on separate sheets, with their desired position in the text clearly indicated. Symbolic formulae appearing as separate lines should be clearly distinguished in the text, and foreign letters (Greek, German, etc.) indicated in the margin. Material to be printed in italics, especially English letters used as symbols ('*A*', '*x*', etc.), should be underlined. On matters of style and layout, authors should consult CHAUNDY, T. W. *et al.* [1964]: *The Printing of Mathematics*, London: Oxford University Press.

Offprints of articles may be ordered when corrected galley proofs are returned. Fifty are supplied free ; the price of extra copies is given on the order form.

© British Society for the Philosophy of Science 1976

For permission to reproduce material from *The British Journal for the Philosophy of Science*, please apply to the Editors.

Back Volumes

Volumes 1-16 may be obtained from Kraus Reprint Ltd., FL-9491 Nendeln, Liechtenstein or Kraus Reprint Corporation, 16 East 46 Street, New York, N.Y. 10017. Volumes 17-25 may be obtained from Aberdeen University Press, Farmers Hall, Aberdeen, Scotland AB9 2XT. Prices on application.

CUL-H00667-18-P8584

Brit. J. Phil. Sci. Vol. 27 No. 1 March 1976

501'05
B774

The British Journal for the Philosophy of Science

Contents

Articles

1–23 ADOLF GRÜNBAUM Can a Theory Answer more Questions than one of its Rivals?
25–42 PAVEL TICHÝ Verisimilitude Redefined.
43–50 ALFRED LANDÉ The Laws behind the Quantum Laws.

Discussions

51–60 SPIRO LATSIS The Limitations of Single-Exit Models: Reply to Machlup.
60–65 MICHAEL DEAKIN On Urbach's Analysis of the 'I.Q. Debate'.
65–73 ROLAND PUCCETTI The Mute Self: A Reaction to DeWitt's Alternative Account of the Split-Brain Data.
73–78 DANIEL ROBINSON What Sort of Persons are Hemispheres? Another Look at 'Split-Brain' Man.

Review Article

79–92 GREGORY CURRIE Was Frege a Linguistic Philosopher? (review of M. Dummett: *Frege: The Philosophy of Language*).

Review

93–96 SUSAN HAACK on W. V. Quine: *The Roots of Reference.*

Recent Publications

Programme of Meetings

Announcement

11 MAY 1976

Printed in Great Britain at the University Press, Aberdeen

THE BRITISH SOCIETY FOR THE PHILOSOPHY OF SCIENCE

Committee 1975-6

President: Professor D. J. O'Connor
Vice-President: Professor J. W. N. Watkins
Hon. Secretary: Helena Sheiham
Hon. Treasurer: Mr. D. Miller
Members: Mr. C. Howson, Dr M. Machover, Mr H. Martins, Mr. A. Young

EDITORIAL BOARD OF THE JOURNAL

Professor R. B. Braithwaite
Professor H. Dingle
Professor Sir John Eccles
Professor Sir Harold Jeffreys
Professor W. C. Kneale
Professor S. Kőrner
Professor M. Polanyi
Professor Sir Karl Popper
Professor G. J. Whitrow
Professor J. H. Woodger

The British Journal for the Philosophy of Science is the official organ of the Society, whose purpose is described by the following statement from its Constitution and Rules :

' The purpose of the Society is to study the logic, the method and the philosophy of science as well as those of the various special sciences, including the social sciences.'

The Society holds meetings, at which papers are read and discussed, once a month during the academic year. These are held, by kind permission, in University College, London. The Committee may recognise Regional Groups of Members formed to hold meetings in other parts of the country. Any Regional Group so recognised may nominate a representative Member to the Committee.

Ordinary membership is open to all persons who have been proposed by two members, approved by the Committee, and elected at an ordinary meeting. Membership of the Society is international and is by no means confined to people resident in the United Kingdom. The annual subscription is £1·00 ($2.50). Those who wish to join should write to the Hon. Secretary, Helena Sheiham, Department of Philosophy, London School of Economics, Houghton Street, London WC2A 2AE.

Brit. J. Phil. Sci. Vol. 27 No. 2 June 1976

The British Journal for the Philosophy of Science

Contents

Articles

105–136 ADOLF GRÜNBAUM Is the Method of Bold Conjectures and Attempted Refutations *Justifiably* the Method of Science?
137–146 DAVID PAPINEAU Ideal Types and Empirical Theories

Discussions

147–159 KARL POPPER A Note on Verisimilitude
160 JON DORLING Comments on Landé
160–161 JON DORLING & DOROTHY EDGINGTON The Applicability of Bayesian Convergence-of-Opinion Theorems to the Case of Actual Scientific Inference
161–164 T. M. BENDITT & DAVID ROSS Newcomb's 'Paradox'

Review Articles

165–176 PETER CLARK Elkana on Helmholtz and the Conservation of Energy (review of Y. Elkana: *The Discovery of the Conservation of Energy*.)
177–185 YORICK WILKS Dreyfus's Disproofs (review of H. Dreyfus: *What Computers Can't Do*.)

Reviews

187–197 JOHN BELL on F. R. Drake: *Set Theory: An Introduction to Large Cardinals* and T. J. Jech: *The Axiom of Choice*; ILHAM DILMAN on R. Wollheim (*ed.*): *Freud, A Collection of Critical Essays*

Announcements

14 SEP 1976

Printed in Great Britain at the University Press, Aberdeen

THE BRITISH SOCIETY FOR THE PHILOSOPHY OF SCIENCE

Committee 1976-7

President: Professor D. J. O'Connor
President-Elect: Mr. L. J. Cohen
Hon. Secretary: Helena Sheiham
Hon. Treasurer: Mr. D. W. Miller
Members: Mr. C. Howson, Dr M. Machover, Mr H. Martins, Mr. A. Young

EDITORIAL BOARD OF THE JOURNAL

R. B. Braithwaite
H. Dingle
Sir John Eccles
M. B. Hesse
Sir Harold Jeffreys
W. C. Kneale
S. Körner
J. L. Mackie
Sir Karl Popper
H. R. Post
J. W. N. Watkins
G. J. Whitrow
J. H. Woodger

The British Journal for the Philosophy of Science is the official organ of the Society, whose purpose is described by the following statement from its Constitution and Rules :

' The purpose of the Society is to study the logic, the method and the philosophy of science as well as those of the various special sciences, including the social sciences.'

The Society holds meetings, at which papers are read and discussed, once a month during the academic year. These are held, by kind permission, in University College, London. The Committee may recognise Regional Groups of Members formed to hold meetings in other parts of the country. Any Regional Group so recognised may nominate a representative Member to the Committee.

Ordinary membership is open to all persons who have been proposed by two members, approved by the Committee, and elected at an ordinary meeting. Membership of the Society is international and is by no means confined to people resident in the United Kingdom. The annual subscription is £1·00 ($2.50). Those who wish to join should write to the Hon. Secretary, Helena Sheiham, Department of Philosophy, London School of Economics, Houghton Street, London WC2A 2AE.

Brit. J. Phil. Sci. Vol. 27 No. 3 September 1976

The British Journal for the Philosophy of Science

Contents

Articles

201–223 IMRE LAKATOS A Renaissance of Empiricism in the Recent Philosophy of Mathematics?
225–229 MENDEL SACHS On the Logical Status of Equivalence Principles in General Relativity Theory
231–249 SUSAN HAACK The Pragmatist Theory of Truth

Discussions

251–258 JACK TIZARD Progress and Degeneration in the 'I.Q. Debate'
259–261 STEVE WYKSTRA On Einstein's Second Postulate
261–267 M. F. PODLAHA Some New Aspects of Relativity: Remarks on Zahar's Paper
267–274 G. SCHLESINGER Unpredictability: A Reply to Cargile and to Benditt and Ross

Review Article

275–287 F. B. D'AGOSTINO Rethinking Transformational Linguistics (review of B. Derwing: *Transformational Grammar as a Theory of Language Acquisition* and G. Sampson: *The Form of Language*.)

Reviews

289–302 I. J. GOOD on R. Swinburne: *An Introduction to Confirmation Theory;* DAVID MILLER on D. A. Gillies: *An Objective Theory of Probability;* JON DORLING on J. Bub: *The Interpretation of Quantum Mechanics,* R. S. Cohen and M. Wartofsky (*eds.*): *Logical and Epistemological Studies in Contemporary Physics* and C. A. Hooker (*ed.*): *Contemporary Research in the Foundations and Philosophy of Quantum Mechanics.*

303–306 *Recent Publications*

Printed in Great Britain at the University Press, Aberdeen

12 NC 1976

THE BRITISH SOCIETY FOR THE PHILOSOPHY OF SCIENCE

Committee 1976-7

President: Professor D. J. O'Connor
President-Elect: Mr. L. J. Cohen
Hon. Secretary: Helena Sheiham
Hon. Treasurer: Mr. D. W. Miller
Members: Mr C. Howson, Dr M. Machover, Mr H. Martins, Mr A. Young

EDITORIAL BOARD OF THE JOURNAL

R. B. Braithwaite
H. Dingle
Sir John Eccles
M. B. Hesse
Sir Harold Jeffreys
W. C. Kneale
S. Körner
J. L. Mackie
Sir Karl Popper
H. R. Post
J. W. N. Watkins
G. J. Whitrow
J. H. Woodger

The British Journal for the Philosophy of Science is the official organ of the Society, whose purpose is described by the following statement from its Constitution and Rules :

' The purpose of the Society is to study the logic, the method and the philosophy of science as well as those of the various special sciences, including the social sciences.'

The Society holds meetings, at which papers are read and discussed, once a month during the academic year. These are held, by kind permission, in University College, London. The Committee may recognise Regional Groups of Members formed to hold meetings in other parts of the country. Any Regional Group so recognised may nominate a representative Member to the Committee.

Ordinary membership is open to all persons who have been proposed by two members, approved by the Committee, and elected at an ordinary meeting. Membership of the Society is international and is by no means confined to people resident in the United Kingdom. The annual subscription is £1·00 ($2.50). Those who wish to join should write to the Hon. Secretary, Helena Sheiham, Department of Philosophy, London School of Economics, Houghton Street, London WC2A 2AE.

Brit. J. Phil. Sci. Vol. 27 No. 4 December 1976

The British Journal for the Philosophy of Science

Contents

Articles

307–315 **Aidan Sudbury** Scientific Laws that are neither Deterministic nor Probabilistic
317–328 **Erol Giray** An Integrated Biological Approach to the Species Problem
329–362 **Adolf Grünbaum** *Ad Hoc* Auxiliary Hypotheses and Falsificationism

Discussions

363–381 **David Miller** Verisimilitude Redeflated
381–391 **Paul Feyerabend** Logic, Literacy and Professor Gellner
392–395 **John Price** On the Probability of Laws being Zero
395–397 **Kenneth Manders** Friedman's Criterion for Simplicity
398 **Michael Bradie** On Writing Popperian History
399–402 **Jarrett Leplin** Synchronisation Rules and Transitivity

Reviews

403–414 **Brian Carr** on D. Lewis *Counterfactuals*; **Noretta Koertge** on J. Agassi *Science in Flux*; **G. E. R. Lloyd** on G. Vlastos *Plato's Universe*; **Richard Swinburne** on E. Sober *Simplicity*.

17 MAR 1977

Printed in Great Britain at the University Press, Aberdeen

THE BRITISH SOCIETY FOR THE PHILOSOPHY OF SCIENCE

Committee 1976-7

President: Professor D. J. O'Connor
President-Elect: Mr. L. J. Cohen
Hon. Secretary: Helena Sheiham
Hon. Treasurer: Mr. D. W. Miller
Members: Mr C. Howson, Dr M. Machover, Mr H. Martins, Mr A. Young

EDITORIAL BOARD OF THE JOURNAL

R. B. Braithwaite
H. Dingle
Sir John Eccles
M. B. Hesse
Sir Harold Jeffreys
W. C. Kneale
S. Körner
J. L. Mackie
Sir Karl Popper
H. R. Post
J. W. N. Watkins
G. J. Whitrow
J. H. Woodger

P, 6584

The British Journal for the Philosophy of Science is the official organ of the Society, whose purpose is described by the following statement from its Constitution and Rules :

' The purpose of the Society is to study the logic, the method and the philosophy of science as well as those of the various special sciences, including the social sciences.'

The Society holds meetings, at which papers are read and discussed, once a month during the academic year. These are held, by kind permission, in University College, London. The Committee may recognise Regional Groups of Members formed to hold meetings in other parts of the country. Any Regional Group so recognised may nominate a representative Member to the Committee.

Ordinary membership is open to all persons who have been proposed by two members, approved by the Committee, and elected at an ordinary meeting. Membership of the Society is international and is by no means confined to people resident in the United Kingdom. The annual subscription is £1·00 ($2.50). Those who wish to join should write to the Hon. Secretary, Helena Sheiham, Department of Philosophy, London School of Economics, Houghton Street, London WC2A 2AE.

Brit. J. Phil. Sci. 27 (1976), 1–23 *Printed in Great Britain*

Can a Theory Answer more Questions than one of its Rivals?*

by ADOLF GRÜNBAUM

1 *Introduction.*
2 *Some Important Ramifications of Thesis* Q.
3 *Is* $QC(N) \subset QC(E)$?
4 *Conclusion.*

1 INTRODUCTION

Karl Popper ([1972], pp. 52–3) has adduced Einstein's and Newton's rival theories of gravitation as a poignant instance of the following claim made by him: If a non-metrical notion of the content of a theory T is construed as the set of all those *questions* to which T can provide *answers*, then the contents of at least some logically *incompatible* extant scientific theories are qualitatively comparable with respect to the relation of proper inclusion.

Several important reasons prompt me to inquire whether Popper has been able to sustain this thesis '*Q*' as to the erotetic comparability of rival theories in empirical science:

(*i*) Thesis *Q* avowedly plays a cardinal role, as we shall see, as one of the pillars of Popper's own account of 'the logical basis of the method of science' ([1972], p. 53).

(*ii*) Popper seems to have overlooked that rank-order comparability of *rival* theories with respect to a kind of *non*-metrical content *other than Tarski's* is urgently needed by him, if his major criterion for the admissibility of theory-modification by auxiliary hypotheses is to have any hope of being applicable even to one single case!

(*iii*) Non-Popperian accounts of scientific endeavour as a problem-solving activity may likewise invoke some variant of Thesis *Q*—if not *Q* itself—to articulate the *rationality* which is presumed to animate certain historic cases of theory-*replacement* within empirical science.

(*iv*) Some challenging but neglected tasks for erotetic logicians will emerge from my inquiry into the erotetic comparability of Einstein's theory of gravitation—hereafter denoted with Popper by '*E*'—with Newton's, hereafter denoted by '*N*'.

Received 24 *October* 1975

---

* The author thanks the National Science Foundation (U.S.A.) for the support of research.

Before turning to the latter inquiry, let me elaborate on those reasons for it which are stated under (*i*) and (*ii*) above.

2 SOME IMPORTANT RAMIFICATIONS OF THESIS Q

(*i*) Let us denote the ***Tarskian logical content*** of a theory T by '$LC(T)$'. Obviously the Tarskian contents of two theories are comparable with respect to set inclusion if one of them entails the other. But equally clearly, if T_1 and T_2 are logically incompatible, then $LC(T_1) \nsubseteq LC(T_2)$ and also $LC(T_2) \not\subset LC(T_1)$. For their incompatibility assures that each of their logical contents has at least one member that does not belong to the other. Hence logically incompatible theories such as E and N have ***Tarskian*** contents which are ***not*** comparable with respect to the subset relation. In view of the latter incomparability, Popper wishes to generalise Tarski's concept $LC(T)$ so as to achieve generalised qualitative content-comparability for at least ***some*** rival theories as well. But just why does Popper wish to achieve generalised non-metrical content-comparability for such rival theories as E and N? As is very clear from his statement ([1972], pp. 52–3), his principal logical motivation springs from the following two sets of considerations:

(*a*) For a pair of theories whose ***Tarskian logical contents*** and falsity-contents are comparable with respect to the (proper) subset relation, Popper can prove the proposition that 'the stronger theory, the theory with the greater content, will also be the one with the greater [non-metrical] verisimilitude ***unless its falsity content is also greater***' ([1972], p. 53). Thus, the latter proposition 'P' is provable for the case of theories one of which unilaterally ***entails*** the other. But Popper ***thinks*** that by ***generalising*** the notion of the non-metrical content of a theory, he can validly extend the purview of Prop. P to any two ***incompatible*** theories whose ***generalised*** contents do stand in the required relation of proper inclusion (or in the required superset relation) and whose falsity contents are comparable. Thus, Popper believes that his generalisation of Tarski's non-metrical notion of content has demonstrably brought at least some pairs of incompatible major theories within the scope of the warranted applicability of Prop. P, when the latter is construed in the ***stronger*** sense of pertaining to ***generalised*** content.

(*b*) Given the latter belief, he then deduces the following: If a theory B, such as E, does exceed a contrary theory A, such as N, in ***generalised*** content, and if the results of attempted falsifications of the 'richer' theory B were to furnish grounds for the conjecture that $B_F \subseteq A_F$, then the richer theory B would presumably be qualitatively more verisimilitudinous than A. And having rightly deemed Prop. P validly applicable to a pair of theories one of which exceeds the other in *Tarskian logical content* by *unilaterally entailing* it, Popper ([1972], p. 53) then in turn draws the following cardinal methodological conclusion: 'The method of bold conjectures and attempted refutations' is *justifiably* 'the method of science', because the quest for falsifications of theories of qualitatively greater *generalised* content has the capability of conducing to the survival of theories of greater verisimilitude in the scientific community.

In a subsequent issue of this *Journal*, I shall deal critically, among other things, with aspects of Popper's theory of verisimilitude relevant to the part of his reasoning just summarised under (*b*). And in my contribution to the forthcoming Lakatos Memorial Volume ([1976]), I have examined in detail several other major pillars of his falsificationist conception of science.[1] Here my concern will be with the role and credentials of Thesis *Q*. Hence I shall now confine myself to commenting on the motivation stated under (*a*) above for Popper's generalisation of the notion of non-metrical content. And I shall now argue that even if Popper could sustain Thesis *Q* by reference to *E* and *N*—as he attempts to do ([1972], pp. 52–3)—his alleged extension of Prop. *P* to *incompatible* theories would still be a *non-sequitur*.

Let us see first how Popper does validly prove Prop. *P* for the case of compatible theories one of which unilaterally entails the other.

In his contribution to the Feigl *Festschrift* ([1966], p. 350, Theorem 1), Popper assumed that the class T of all true statements is not finitely axiomatisable and considered finitely axiomatisable theories A and B for which he then proved the following 'theorem on truth-content' (*cf.* also [1972], p. 52, n. 20):

---

[1] Some of my central contentions in that essay are the following: (1) Popper's historiography of inductivism is one-sidedly oblivious of a long tradition of inductivism, beginning with Bacon, which *denies* that every instance of a hypothesis is *automatically* supportive of the hypothesis to some non-zero degree or other, and *a fortiori* denies that all instances of a hypothesis count as equally supportive to some non-zero degree. And thus Popper overlooks that the mere quest for credibilifying a hypothesis by supporting evidence does not commit inductivism as such to either of the stated forms of the instantiation condition. As a result, the test case of psychoanalysis is shown not to illustrate the superiority of Popper's falsificationist criterion of demarcation to inductivist requirements for scientific status. (2) Popper's claim that both the *a priori* and *a posteriori* probabilities of *any* universal law-statement must inevitably be *zero* contradicts his *quantitative* theory of the content and verisimilitude of a hypothesis, which asserts that some universal theories *differ* in both measured content and *a priori* probability. Hence, if his quantitative theory of content and verisimilitude are to be upheld, he cannot adduce the vanishing probabilities of all universal theories as a *reductio ad absurdum* argument against probabilistic inductivism. (3) Popper cannot supply a purely deductivistic rationale for advocating *severe* tests of a hypothesis or for having successful *risky* predictions *count in favour* of the hypothesis. For there is no rational basis on which a pure deductivist can even hope for a good chance that *severe* tests will weed out false theories any better than *non*-severe tests. (4) There is a major inconsistency in Popper's *1972* account of corroboration. (5) Detailed comparison of Popper's theory of corroboration with (non-subjectivistic) Bayesian inductivism provides further evidence that Popper has not furnished a viable alternative to all forms of inductivism on genuinely deductivist foundations. (6) Falsifiability is not the touchstone of scientific rationality *to the exclusion* of inductive supportability.

$$[LC(A) \subset LC(B)] \leftrightarrow A_T \subset B_T,$$

where 'A_T' and 'B_T' denote the respective *truth*-contents of the theories A and B.[2] (Similarly, below we shall denote Popper's falsity-content of a theory A by 'A_F').

Popper's qualitative definition of lesser verisimilitude ([1962], p. 233; [1972], p. 52) as rendered by Pavel Tichy ([1974], p. 156, Prop. 2.3) is as follows:

(*Def.*) 'A *has less verisimilitude than* B *just in case either* $A_T \subset B_T$ *and* $B_F \subseteq A_F$ *or* $A_T \subseteq B_T$ *and* $B_F \subset A_F$'.

We shall write '$A <_{vs.} B$' as short for 'A has less qualitative verisimilitude than B'. Since this definition gives a necessary and sufficient condition for $A <_{vs.} B$, we see that the following condition is *sufficient*:

$$A_T \subset B_T \quad \text{and} \quad B_F \subseteq A_F.$$

In virtue of the transitivity of entailment *and* the exportation rule, the conjunction of the latter sufficient condition for $A <_{vs.} B$ with Popper's above theorem on truth content enables Popper to deduce that

$$\text{(Prop. } P\text{)} \quad \{LC(A) \subset LC(B) \quad \text{and} \quad B_F \subseteq A_F\} \rightarrow A <_{vs.} B.$$

Now let us employ the symbol '$QC(T)$' to denote the (unmeasured) set of questions *answered* by the theory T, an acronym which abbreviates the term 'The Question-Content of T'. Could Popper proceed in the manner of his deduction of Prop. P by invoking his theorem on truth-content to assert that $[QC(N) \subset QC(E)] \rightarrow N_T \subset E_T$ and thence infer, *via* the stated sufficient condition for $A <_{vs.} B$, that—for *incompatible* theories like N and E—

$$\text{(Prop. } P^*\text{)} \quad \{QC(N) \subset QC(E) \quad \text{and} \quad E_F \subseteq N_F\} \rightarrow N <_{vs.} E?$$

To see that the answer is plainly negative, note that the theorem on truth-content is proven only for Tarskian logical content and *not* for

---

[2] David Miller ([1974*a*] and [1974*b*], pp. 168, 168 n. 1, and 174, Corollary 2 of Theorem 5) modified Popper's assumptions and then carried out the proof of the corresponding theorem of truth-content as follows: (*a*) The class T is now assumed *not* to be even recursively axiomatisable, which is a stronger, more restrictive (*less* general) property than the property of not being finitely axiomatisable that Popper had assumed. (*b*) Miller also assumes that the two theories A and B whose content ranking is to be proven to be the same as their truth-content ranking are recursively but not necessarily finitely axiomatisable, whereas Popper had assumed more restrictively that A and B are finitely axiomatisable. Thus, by then proving that an increase in truth-content as between two *recursively axiomatisable* systems (which Miller calls 'axiomatisable') goes hand-in-hand with an increase as between their contents, Miller has *generalised* the result of Popper's 1966 proof of the corresponding theorem for *finitely axiomatisable* systems (which Popper calls 'axiomatisable'). In short, by thus covering a wider, less restrictive class of theories A and B in the theorem, Miller has strengthened Popper's result.

question-content QC. Clearly therefore, Popper's premiss $QC(N) \subset QC(E)$ can hardly serve to vouchsafe *via* his theorem on truth-content that $N_T \subset E_T$. And the latter claim is thus unavailable as a link in the deduction of Prop. P^*.

Moreover, suppose that *in order to deduce Prop. P^* via* the stated sufficient condition for $N <_{vs.} E$, Popper *were* to assert the entailment $[QC(N) \subset QC(E)] \rightarrow (N_T \subset E_T)$. Given his espousal of $QC(N) \subset QC(E)$, the latter assertion would then yield the demonstrably *false* conclusion $N_T \subset E_T$. That this conclusion is indeed *false* follows from the premiss $LC(N) \not\subset LC(E)$—which is guaranteed by the logical incompatibility of N and E—in conjunction with the theorem on truth-content. Thus, Popper could not consistently invoke the stated assertion of entailment in order to derive Prop. P^*.

Hence, unlike Prop. P, its counterpart Prop. P^* is *at least unfounded*. Yet Popper invokes Prop. P^* and the premiss $QC(N) \subset QC(E)$ to claim ([1972], p. 53) that if $E_F \subseteq N_F$, then $N <_{vs.} E$.

(ii) In his LSD ([1959], section 20, pp. 82–3) Popper writes:

> As regards *auxiliary hypotheses* [introduced to accommodate instances which refute the *original* form of the theory] we decide to lay down the rule that only those are acceptable whose introduction does not diminish the degree of falsifiability or testability of the system in question, but, on the contrary, increases it. (How degrees of falsifiability are to be estimated will be explained in sections 31 to 40.)

Upon turning to his sections 31–40 (pp. 112–35), we find him telling us (in sections 32, 33 and 38) that the degrees of falsifiability of two theories may be comparable not only *via* subset relations among their content-classes but also by means of their '*dimensions*', *i.e.* by means of 'the degree of complexity of their simplest falsifiers' ([1974], Book II, p. 1047). And he then adds significantly ([1959], p. 130) that there will be cases in which neither or only one of the two methods of comparison of degrees of falsifiability is applicable. Thus, he does not tell us what to do when neither is applicable. In any case, it is clear that in his LSD, he intends the achievement of *increase of content* to qualify as *sufficient* to legitimate the postulation of a modifying auxiliary hypothesis without asserting that it is also *necessary*, although he does require an increase of the *degree of falsifiability* as certified by at least one of his two criteria. Yet in his 'Intellectual Autobiography', he seems to be saying that *increase of content* is both necessary and sufficient to legitimate a modifying auxiliary. For there he writes:

> in *Logik der Forschung* . . . I introduced *degrees of testability*, and these turned out to be closely related to (degrees of) *content*, and surprisingly fertile: increase

of content became the criterion for whether we should, or should not, tentatively adopt an auxiliary hypothesis ([1974], p. 32).

When discussing the 'immunisation' of a theory from a falsification by the legitimate introduction of a modifying auxiliary, Popper ([1974], Book I, p. 32) cites the paradigm example of the postulation of the planet Neptune, which served to explain observations of the motion of Uranus that had been incompatible with the prior version of Newtonian planetary theory. As is apparent from his account there of this example and from his illustrative mention ([1959], p. 83) of other auxiliary hypotheses, his stated requirement for the admissibility of an auxiliary is typically addressed to the following kind of logical situation: Existing evidence E is logically incompatible with a theory T_1, and a theory T_2 generated from T_1 by a modifying auxiliary hypothesis H is then able to explain E.

Here I shall not inquire whether in any of the known cases of theory-rescue by a modifying auxiliary H, Popper's *dimension method* is capable of assessing compliance with his stated LSD requirement for the admissibility of H. Instead, I am concerned to demonstrate the following: If the non-metrical content of a theory T is construed as either the Tarskian logical content $LC(T)$ or as Popper's informative content, or yet as Popper's empirical content (which is a species of his informative content), then no auxiliary H can EVER generate a *rescuing* T_2 of *greater content*! And thus the desideratum that there be an increase of *any such* non-metrical content is a wholly *useless* and *vacuous* sufficient condition for compliance with Popper's rule of admissibility. Furthermore, if *such* content-increase were also held to be a necessary condition for the admissibility of H, then *no* modifying auxiliary could *ever* satisfy it.

Once I have demonstrated this below, it will be seen to be imperative to inquire whether the construal of the non-metrical content of a theory T as its *question*-content $QC(T)$ can make Popper's criterion of content-increase non-vacuously applicable to the admissibility of rescuing auxiliaries after all. Hence, as a corollary of my examination of Thesis Q in 3, I shall comment on whether the postulation of Neptune can be held to comply with Popper's rule of admissibility in virtue of having generated a planetary theory of increased *question*-content.

Turning to my demonstration of the stated vacuity of Popper's specified sufficient condition for admissibility, consider quite generally *any* case in which a *modus tollens* argument against an original theory T_1 on the basis of *refuting* evidence E prompts the modification of T_1 by means of a *new auxiliary hypothesis*, so as to issue in a theory T_2 which then does *explain* (entail) E. Clearly, since T_1 entails the *negation* of E while T_2 entails E, the theories T_1 and T_2 are *logically incompatible*. But, as we already had occasion to note, in the case of any two incompatible theories whatever, the consequence class of *neither* theory can *ever* be a subset of the other. Therefore, the *LC* of our *rescuing* T_2 can *never* properly include the *LC* of its incompatible superseded ancestor T_1. And hence to establish my aforestated contention of vacuity, all that we still need to show is that the same *incomparability* of non-metrical content holds for Popper's notions of 'informative content' and 'empirical content'.

He defines the 'informative content' of a theory T—which I shall denote by '$IC(T)$'—as follows: '*the informative content of a theory is the set of statements which are incompatible with the theory*' ([1974], Book I, p. 18). And in an appended footnote ([1974], Book I, p. 158, n. 16) he recalls his LSD definition of 'empirical content'—which I shall denote by '*EC*'—as follows: 'The subset of the informative content which consists of basic statements (empirical statements) I call in *L. Sc. D.* the class of the theory's "potential falsifiers", or its "empirical content".' Thus, despite the labels 'informative content' and 'empirical content', every statement belonging to either $IC(T)$ or, more restrictedly, to $EC(T)$ is logically a *contrary* of T. And though *prima facie* incongruous, these labels do cohere with Spinoza's dictum '*Omnis affirmatio est negatio*'. Indeed Popper justifies his use of the term 'informative content' by making the following observation concerning the set $IC(T)$:

> It may be seen at once, however, that the elements of this set and the elements of the logical content stand in a one-one correspondence: to every element which is in one of the sets, there is in the other set a corresponding element, namely its *negation*.
>
> We therefore find that if the logical strength, or the power, or the amount of information in a theory increases or decreases, its logical content and its informative content must both likewise increase or decrease. This shows that the two ideas are very similar: there is a one-one correspondence between what can be said about the one, and what can be said about the other. Thus my definition of informative content is not entirely absurd ([1974], Book I, p. 18).

Mr Frederick W. Sawyer, III has pointed out in effect that if Popper's formulation here of the relation between $LC(T)$ and $IC(T)$ is to be sound, the term 'statement' has to be understood in the extra-linguistic sense of proposition rather than in the linguistic sense of *sentence*.[1]

---

[1] Sawyer notes that *if* statements were to be *sentences*, then the reason given here by Popper for claiming a one-to-one correspondence under negation between the statement-elements of $LC(T)$ and those of $IC(T)$ will be *insufficient*: The correct assertion that 'to every element which is in one of the sets, there is in the other set a corresponding element, namely its *negation*' does not suffice to guarantee the one-to-one correspondence claimed by Popper, unless the following tacit assumptions are made: Not only $LC(T)$ but also $IC(T)$ is closed under logical equivalence, *and* the arguments as well as values of the negation function are not sentences but rather equivalence classes of sentences with respect to the relation of *logical* equivalence.

In this last quotation from him, Popper rightly pointed out that

$$(T_1)(T_2)\quad [\{IC(T_1) \subset IC(T_2)\} \leftrightarrow \{LC(T_1) \subset LC(T_2)\}]$$

Now apply this result to the special case of an original theory T_1 and a *contrary rescuing* theory T_2 generated from T_1 by an auxiliary H. Since T_1 and T_2 are incompatible, $LC(T_1) \not\subset LC(T_2)$. And it then follows from Popper's result that the *informative* content $IC(T_1)$ of T_1 can *never* be a proper subset of $IC(T_2)$. Indeed, the stronger conclusion that $IC(T_1) \nsubseteq IC(T_2)$ can be seen to follow at once from the mere fact that, by hypothesis, $E \in IC(T_1)$ while $E \notin IC(T_2)$. Recall Popper's 'Spinozistic' definition of the empirical content $EC(T)$ of T as that subset of $IC(T)$ [*not* of $LC(T)$!] which consists of 'basic' statements. Since E is 'basic', it also follows that $E \in EC(T_1)$ while $E \notin EC(T_2)$. Hence $EC(T_1) \nsubseteq EC(T_2)$. And this completes the demonstration of my vacuity contention.

Now let τ_1 be the *pre*-Neptunian version of Newtonian planetary theory of *circa* 1820, while τ_2 is the amended Adams and Leverrier version of solar planetary theory which asserts the existence of the additional trans-Uranian planet Neptune. Then the auxiliary hypothesis H is the assertion that this additional planet exists *and* exerts non-negligible gravitational influences on the motion of Uranus. Furthermore, if U_1 is the 'basic' statement asserting τ_1's *incorrect* prediction of the orbit of Uranus, and U_2 is the *contrary* basic statement codifying the actually 'observed' orbit of Uranus, then τ_2 entails U_2, which thus plays the role of our paradigmatic E above. And since τ_1 and τ_2 are just a special case of our incompatible ancestor and rescuing theories T_1 and T_2, we have at once that

$$IC(\tau_1) \not\subset IC(\tau_2) \quad \text{and} \quad EC(\tau_1) \not\subset EC(\tau_2).$$

It is very surprising that Popper *explicitly denies* both of these patent results when he writes as follows:

> a prima facie falsification *may* be evaded ... as in the Uranus/Neptune sort of case, by the introduction of testable auxiliary hypotheses, so that the empirical content of the system—consisting of the original theory plus the auxiliary hypothesis—is greater than that of the original system. We may regard this as an increase of informative content... There are, of course, also auxiliary hypotheses which are merely evasive immunising moves. They decrease the content ([1974], Book I, p. 33, item (*e*)).

It will be noted that here Popper asserts both $EC(\tau_1) \subset EC(\tau_2)$ and $IC(\tau_1) \subset IC(\tau_2)$. To see how he was led to these false conclusions, observe that he uses the word '*plus*' when he characterises the theoretical system τ_2 as 'consisting of the original theory [τ_1] plus the auxiliary hypothesis [H]'. His use of the word '*plus*' here seems to have misled him as follows: He logically misconstrued the so-called 'introduction' of the hypothesis H of an additional planet as a *mere conjunctive appending* of H to τ_1, thereby inferring that τ_2 is the conjunction of τ_1 and *H*! In this way, he was apparently led to make the untenable assertion that the transition from τ_1 to τ_2 involved an *increase* of empirical and hence informative content. But this claim runs afoul of the following facts: (1) Far from generating τ_2 by being *appended conjunctively* to τ_1, the full-blown auxiliary *H replaces* the *contrary* constituent of τ_1 which (tacitly) asserts the *non*-existence of any trans-Uranian planet capable of exerting appreciable gravitational forces on Uranus, (2) τ_1 and τ_2 are logically incompatible, if only because they respectively entail the contrary 'basic' orbital statements U_1 and U_2. Having apparently misdepicted τ_2 as the conjunction $\tau_1 \cdot H$, it is not surprising that Popper feels entitled to claim an increase of *IC* and *EC*.

No wonder, therefore, that Popper overlooked the *vacuity* of any increase in *LC*, *IC* or *EC qua sufficient condition* for the admissibility of an auxiliary.

We can now turn to the principal inquiry of this essay.

3 IS $QC(N) \subset QC(E)$?

Popper addresses himself to the comparability of '*competing* theories—such as the Newtonian and Einsteinian theories of gravitation— . . . with respect to their (unmeasured) contents' ([1972], p. 52) amid noting that 'there are also competing theories which are not comparable' (p. 52). Before seeing just how Popper proposes to use the comparative question-answering capabilities of two competing theories as a criterion of comparison of their contents, let us note his remark concerning this new criterion as a *generalisation* of his earlier Tarskian account of content comparison as follows:

> In the first instance, Tarski's notion of an (unmeasured) consequence class or content permits us to compare the contents of theories if one of them entails the other. The generalisation here given allows us to compare contents (or the measure of contents) if one of them can answer all questions that can be answered by the other, and with at least equal precision ([1972], p. 53, n. 21).

He explains by reference to the example of Newton's and Einstein's theories how he conceives the new criterion to be at once a criterion of comparative content, explanatory power and degree of falsifiability, writing:

> The intuitive comparability of the contents of Newton's theory (*N*) and Einstein's (*E*) can be established as follows [n. 20 omitted here]: (a) to every question to which Newton's theory has an answer, Einstein's theory has an answer which is at least as precise; this makes (the measure of) the content, in a slightly wider sense than Tarski's [n. 21 omitted here, but cited earlier above], of *N* less than or equal to that of *E*; (b) there are questions to which Einstein's

> theory *E* can give a (non-tautological) answer while Newton's theory *N* does not; this makes the content of *N* definitely smaller than that of *E*.
>
> Thus we can compare intuitively the contents of these two theories, and Einstein's has the greater content. (It can be shown that this intuition is borne out by the content measures *ct*(*N*) and *ct*(*E*).) This makes Einstein's theory *potentially* or *virtually* the better theory; for even before any testing we can say: if true, it has the greater explanatory power. Furthermore, it challenges us to undertake a greater variety of tests . . . (*Ibid.*, pp. 52–3).
>
> These are some of the advantages, existing *even before it has been tested*, of a (logically) stronger theory; that is, of a theory with greater content (*ibid.*, p. 53).
>
> . . . A theory is the bolder the greater its content. It is also the riskier: it is the more probable to start with that it will be false (*ibid.*, p. 53).

I defer discussion of the latter *a priori* probability claim until I discuss *quantitative* verisimilitude in a subsequent issue of this *Journal*. But I note that although Popper does not say so in connection with this statement, it would seem that at least part of his logical inspiration for this new proposed, generalised criterion of comparative content and/or degree of falsifiability derives from his repeated emphasis elsewhere that 'Newton's theory lives on as a limiting case in Einstein's theory of gravitation' ([1974], Book I, p. 160, n. 32). And Popper cites a paper by Peter Havas ([1964], pp. 938–65) as having made this particularly clear by means of a four dimensional (space-time) formulation of Newtonian mechanics and by taking the velocity *c* of light as going to infinity. And again citing the same work by Havas, Popper says:

> I have often asserted that Newton's theory explains (and contains as a first approximation) both Kepler's laws and Galileo's law of falling bodies, while Einstein's theory contains, in a similar way, Newton's.
>
> It is a strange way of 'containing', for Newton's theory logically contradicts Kepler's and Galileo's, just as Einstein's logically contradicts Newton's. The relationships here are very interesting, and deserve closer study (such as the studies by Peter Havas quoted in n. 205) ([1974], Book II, p. 1151).

I shall now offer a series of questions which qualify at least *prima facie* as counter-examples to Popper's claim that 'to every question to which Newton's theory has an answer, Einstein's theory has an answer which is at least as precise'. We saw that Popper offers this claim here as his premiss for saying 'this makes (the measure of) the content, in a slightly wider sense than Tarski's [n. 21 omitted here] of *N* less than or equal to that of *E*'. My impending counterexamples to Popper's premiss will serve to show that *N*'s being a limiting case of *E*, in the sense illuminatingly articulated by Havas, in no way militates against the abundant existence of such counterexamples.

I shall consider five sets of counterexamples under the headings 0–4 inclusive. Independently, David Miller ([1975], Sec. *2*) has given some avowedly only trivial examples to impugn the thesis which I call '*Q*'. And in that context, he also addresses himself to the issue posed in the title of the present essay.

Set 0. In Newton's theory of gravitation, the yes-or-no question (0*a*) 'Is *the* geometry of the three-dimensional spaces in which gravity acts in accord with the laws of motion and gravitation Euclidean?' is well-formed and is answered (affirmatively). For Newton tells us in the *Principia* that the three-dimensional absolute space of his primary inertial system and the three-dimensional relative spaces of his secondary inertial systems are 'the same in figure and magnitude' ([1934] p. 6), and that their geometrical structure also remains the same at all times.

Also, since the *instantaneous* spatial distance between gravitationally interacting mass points as well as the instantaneous acceleration of either of them are the same with respect to all inertial systems, the gravitational equations

$$m_1a_1 = -G\frac{m_1m_2}{r^2}$$

and

$$m_2a_2 = -G\frac{m_1m_2}{r^2}$$

hold alike in the Euclidean 3-spaces of *all* the coordinate systems in which Newton's laws are asserted to hold.

But in the context of Einstein's theory of gravitation, the above question (0a) *is doubly ill-formed.*

To illustrate that, in the context of *E*, the question (0*a*) is doubly ill-formed, consider Schrödinger's treatment of the so-called de Sitter model of the universe in the context of Einstein's theory of gravitation. Schrödinger ([1956], ch. I) shows that in the space-time of this universe, the three-dimensional spatial geometry depends not only on the reference frame associated with the choice of a particular space-time coordinate system in that universe but, in general, also on the time in that coordinate system at which the 3-geometry is to have a determinate character. In particular, Schrödinger exhibits one space-time coordinate system in which the spatial 3-geometry is non-Euclidean and finite but otherwise time-dependent in regard to the degree of its curvature, while exhibiting another frame in which 3-space is Euclidean and infinite at all times, though this space expands in the sense that the distances between its space points increase with time. And hence Schrödinger goes on to make the following statement about the latter frame (*ibid.*, pp. 28–9): 'This frame is interesting because it presents space as infinite and *qua* space, flat [Euclidean], showing that the question [0*b*] asked so often, whether space [the *one* presumed geometry of all (instantaneous) 3-spaces] is really curved and finite, cannot be answered, because the answer depends on the frame.'

It is *crucial* to see that what is being queried in the question (o*b*) which Schrödinger just declared unanswerable in *E* is only the particular *identity* of the *one* kind of geometry which is *presupposed* by (o*b*) to characterise all (instantaneous) physical 3-space alike. By the same token, question (o*a*) is predicated on the same presupposition as (o*b*) and, like (o*b*), asks only for the particular *identity* of the *assumedly unique* physical 3-geometry. An interlocutor can ask a question *Q* which *differs* from (o*a*) by *not* assuming that all (instantaneous) physical 3-spaces have the same geometry and asks instead 'Are all of the geometries of the instantaneous physical 3-spaces of nature Euclidean?' That *Q* and (o*a*) ask different questions is illustrated by the fact that a negative answer to *Q* asserts only that *some* instantaneous 3-spaces are *non*-Euclidean, whereas a negative answer to (o*a*) asserts that *all* of them are. Schrödinger's results show that whereas *E can* answer *Q* negatively, it cannot *consistently* answer (o*a*) either 'yes' or 'no', any more than it consistently answer the question (o*b*). For in *E*'s de Sitter world, some instantaneous 3-geometries are Euclidean while others are non-Euclidean and finite. Hence all of the universal statements 'All 3-geometries are Euclidean,' 'No 3-geometries are Euclidean,' 'All 3-geometries are curved and finite,' and 'No 3-geometries are curved and finite' are *false* in *E*'s de Sitter world. Yet affirmative or negative answers to either (o*a*) or (o*b*) assert one of these universal statements. Apart from the particular role of Euclidean 3-spaces as such in Schrödinger's example, note that questions (o*a*) and (o*b*) alike rest on the presupposition of a common identity of all (instantaneous) 3-geometries, a presupposition which Schrödinger showed to be incompatible with *E* or *false* in *E*. Thus a 'no' answer no less than a 'yes' answer to either of these two questions would be implicity *affirming* this presupposition of sameness of 3-geometries. This is so, since either kind of answer would be making an assertion concerning this purportedly unique 3-geometry.

Thus whatever cognate questions *E* does answer, it does *not* answer (o*a*), whose presupposition belongs to *LC*(*N*), *i.e.*, to the logical content of *N*. Yet *N* clearly does answer (o*a*).

Clark Glymour has remarked that there are also *global* topological questions concerning space and/or space-time which are answered by *N* but not by *E*.

It would seem that question (o*a*) alone poses major difficulties for Popper's claim. But it will be instructive to go on to other kinds of questions which are clearly counterexamples.

Set 1. For expository purposes, it will be useful to consider the generally much more familiar special case (treated on p. 942 of Havas's [1964]) in which the Lorentz transformations of special relativity ['STR'] kinematics 'reduce' to the Galilean transformations of Newtonian kinematics in the limit of the dimensionless ratio v/c going to zero. (The latter limiting process may be conceptualised as $v \rightarrow 0$ in a given system of units, which *is* allowed by the STR, rather than that the limiting velocity c does not have the fixed finite value c after all in a given system of units but goes to infinity instead, which is *not* allowed by the STR). Note that the flat Minkowski space-time of the STR is one particular solution of the vacuum field equations of Einstein's general theory of relativity ['GTR']. In *this* sense, the STR is *part* of Einstein's theory of gravitation!

In simplest form, the Galilean transformations can be written as $x' = x - vt$ and $t' = t$. Consider two events A and B. Subtraction of $x'_A = x_A - vt_A$ from $x'_B = x_B - vt_B$ yields $x'_B - x'_A = x_B - x_A - v(t_B - t_A)$. Since $v \neq 0$, an assumed equality of the two spatial separations $x'_B - x'_A$ and $x_B - x_A$ would entail that $t_B - t_A = 0$ or $t_B = t_A$. The same conclusion is deducible from the Galilean transformations in their most general form.

Thus, the statement that 'All events having a Galileanly-invariant non-zero spatial separation in all inertial frames are simultaneous' belongs to the logical content $LC(N)$ of Newton's theory.

Hence, within the framework of N, we can meaningfully ask the following questions and *can answer them* by means of its Galilean transformations:

(1*a*) Are any two punctal events having a non-zero *spatial separation* which is *invariant* in *all* inertial frames under the Galilean transformations necessarily simultaneous?, and/or

(1*b*) Why is it the case that any two punctal events having a non-zero spatial separation which is invariant in all inertial frames must be simultaneous?

These questions can be meaningfully addressed to the Galilean transformations and can be answered by them, because (*i*) the ontology of Newtonian kinematics countenances the existence of a pair of events having the specified kind of invariant spatial separation, and likewise countenances the existence of non-coinciding events which are simultaneous simpliciter, thereby countenancing the existential presuppositions on which question (1*a*) is predicated, and (*ii*) N gives an *affirmative* answer to question (1*a*) and thereby allows the state of affairs for which question (1*b*) seeks an explanation.

What can be said as to whether E 'has an answer' to questions (1*a*) and (1*b*)? For the moment, let us confine ourselves to (1*a*). Here we must first *distinguish* a relativistic *homologue* of (1*a*) from a possible formulation of (1*a*) which is *equivalent*—by a stated logical criterion of equivalence of questions!—and yet well-formed in the context of *E*. For clearly, a putative equivalent formulation of (1*a*) is a statement of 'the *same* question' in a sense very much more essential to Popper's concerns than a mere relativistic homologue.

In particular, the question posed by (1*a*) is *not* the one posed by the following relativistic *homologue*: 'Do all non-coinciding events possessing the relativistically *self-contradictory* property of having an *invariant* non-zero spatial separation *with respect to the Lorentz* [*not* the Galilean!] *transformations* also possess the relativistically self-contradictory property of being invariantly simultaneous with respect to the Lorentz transformations?' I say that having a non-zero spatial separation which is invariant with respect to the Lorentz transformations is a relativistically self-contradictory property of non-coinciding event-pairs, because in the ontology of the kinematics of the Lorentz transformations, it is *impossible* that there be a pair of events having a non-zero spatial separation which is invariant with respect to all inertial frames. The reason is that in the context of the Lorentz transformations, such invariance of the *three*-dimensional spatial separation of two events A and B would entail the following: their time-difference Δt_{AB} in a given inertial frame I and their time-difference $\Delta t'_{AB}$ in any and *every* other inertial frame I' whose relative velocity v is $0 < v < c$ must satisfy the condition $\Delta t'_{AB} = -\Delta t_{AB}$. But there exist frames I' having the specified velocity v such that $\Delta t'_{AB}$ violates this necessary condition. And this violation occurs regardless of whether the *space-time* (*four*-dimensional) separation between A and B is time-like, null, or space-like. Hence A and B cannot have an invariant spatial separation. Corresponding reasons for self-contradictoriness apply relativistically in the case of non-coinciding punctal events to the property of being invariantly simultaneous with respect to the Lorentz transformations. And hence to ask in (1*a*) whether two events having an invariant spatial separation are *Newtonianly* absolutely simultaneous in the kinematical ontology of the Galilean transformations is *not* to ask whether they possess the relativistically self-contradictory property of invariant simultaneity.

How then could Popper furnish a relativistic *equivalent*, *if such there be*, to question (1*a*) rather than an unhelpful relativistic homologue? The homologue is most unhelpful, because it inquires about the relations between the *empty* extensions of two self-contradictory properties, which are *trivially* identical as a matter of logic alone without physics, whereas the Newtonian question (1*a*) does nothing of the kind. Let us cite Popper's account of his notion of 'the *problem content* of a theory' with a view to assessing its potential for furnishing the sought-after putative relativistic equivalent of (1*a*). But let us point out for later purposes that when he introduces the notion of 'problem content' ([1974], p. 18), he characterises it as 'merely a kind of unification of the other two' ideas of content: logical and informative content. By contrast, we saw that his notion of *question*-content of a theory is characterised by him as a *generalisation* of the Tarskian concept of content.

Popper writes:

> Following a suggestion of Frege's, we may introduce the notion of a yes-or-no problem or, briefly, a y-problem: given any statement a (say, 'Grass is green'), the corresponding y-problem ('Is grass green?') may be denoted by '$y(a)$'. One sees at once that $y(a) = y(non\text{-}a)$: the problem whether grass is green is, *qua* problem, identical with the problem whether grass is not green, even though the two questions are differently formulated, and even though the answer 'Yes' to one of them is equivalent to the answer 'No' to the other.
>
> We can define what I propose to call the problem content of a theory t in either of two equivalent ways: (1) it is the set of all those $y(a)$ for which a is an element of the logical content of t; (2) it is the set of all those $y(a)$ for which a is an element of the informative content of t. Thus the problem content is related to the two other contents in identical ways ([1974], Book I, p. 20).

It would seem clear that since Popper characterises the member-questions of the problem-content as 'yes-or-no' problems, he is claiming that *every* question belonging to the problem content of t does receive an unambiguous yes-or-no answer from t.

In the context of E, could Popper recast the Newtonian yes-or-no question (1*a*) in *provably equivalent* fashion into the kind of yes-or-no neo-Fregean problem format $y(a)$ required by his definition of 'problem content'? Perhaps this could be *attempted* along somewhat the following lines by a conjunction of questions: (*i*) Are physical events related by the Galilean transformations?, (*ii*) if so, are there events *in the world* which have an invariant non-zero spatial separation with respect to these transformations, and are there Newtonianly simultaneous events?, and (*iii*) *if* the answers to these two questions are affirmative, is it the case that if two *physical* events *in the world* do have a Galileanly invariant non-zero spatial separation, then these events are Newtonianly simultaneous? Popper has not told us whether this is how he would try it. I think that E clearly has a 'no' answer to constituent (*i*) of the proposed attempted equivalent recast of (1*a*). But in the case of constituents (*ii*) and (*iii*), E denies the antecedents of the questions and hence can be said to have 'obviated' rather than answered them, to use Belnap's apt terminology. Thus E does not answer (*ii*) or (*iii*) except 'vacuously', as it were.

But in any case, would this conjunction of questions be equivalent to

(1*a*)? If not, can we claim quite generally that the yes-or-no question (1*a*) cannot belong to the problem content of *E*, although it manifestly does belong to the problem content of *N*? Let us consider each of these two issues in turn.

I submit that the interlocutor who poses our question (1*a*) is *not* querying what is being queried in the above constituent questions (*i*) and (*ii*) of the attempted recast of (1*a*). Instead, he is predicating his question (1*a*) on affirmative answers to these constituent questions. And hence unlike constituent question (*iii*), the interlocutor who poses (1*a*) is *assuming* on the basis of *N* that the antecedent of question (*iii*) is satisfied: What (1*a*) is querying is *only* whether one *assumedly non*-empty set is included in another *assumedly non*-empty set! Hence (1*a*) is surely not equivalent to the conjunction of (*i*), (*ii*) and (*iii*).

Moreover, while (1*a*) is querying only whether one *assumedly non*-empty set is included in another *assumedly non*-empty set, *E* asserts that both of these sets are empty. For precisely this reason, *E* cannot give an unambiguous yes-or-no answer to (1*a*) but can only obviate it, whereas *N* does answer it affirmatively. Furthermore, since Popper clearly seems to hold that every member-question of the problem-content of a theory *T* does receive an unambiguous yes-or-no answer from *T*, it follows that (1*a*) cannot belong to *E*'s problem-content, although it does belong to that of *N*.

Thus, in the context of the STR-component of *E*, questions (1*a*) and (1*b*) cannot even arise, let alone be answered with at least the same precision as in *N*. For in *E* these questions are ill-formed by being predicated on *assumptions* which are denied by the STR component of *E* but which *are* members of *LC*(*N*). As the medievals said in the case of such questions, '*cadit quaestio*'.

Set 2. In Newton's dynamics, the velocity v attained by a particle of fixed mass m which is accelerated from an initial state of rest by a constant force F acting for a time t is given by

$$Ft = mv.$$

Suppose that one were to ask *N* for an answer to the question

(2*a*) 'How much time is required by a constant force of magnitude F to accelerate a particle of fixed mass m in an inertial frame up to a speed $2c$, which is twice the numerical value of the usual speed c of light *in vacuo*?'

Clearly, *N* would answer precisely that $t = 2mc/F$. But how would *E* deal with such a question? The STR asserts that the fixity of m when the particle is no longer at rest is a false presupposition of the question (2*a*), as is the latter's presupposition that any superlight velocity such as $2c$

can be attained at all by a particle. Thus in E, the demand of (2*a*) for a time specification t would be ill-conceived, even if t is allowed to go to infinity! How then can E (STR in this case) possibly be held to answer (2*a*) at all, let alone at least as precisely as N does?

By the same token, there are infinitely many questions (2*b*) about the outcomes of, say, collisions of positive mass particles under specified initial conditions involving their having superlight velocities in (local) inertial frames, such that these questions can be *asked* and *answered* in N but not in E, let alone with at least equal precision. For E denies the very physical possibility of the initial conditions on which each one of this infinitude of questions is predicated, whereas N asserts these same initial conditions to be physically realisable. Thus it is utterly unavailing here that the assertion of the possibility of superlight particle velocities in inertial frames belongs to *both* $IC(E)$ and $LC(N)$!

The case of a particular set of initial conditions I *involving the motion of a positive mass particle* A *with a super-light velocity* on a collision course with another such particle B now enables us to show more precisely that there are an infinitude of questions (2*b*) of the form 'What happens under circumstances I?' such that (*i*) these questions (2*b*) are well-conceived in the context of N, while (*ii*) they are ill-conceived in the context of E, and E does not answer but only *obviates* them.

(*i*) If F is the collision-*outcome* predicted by N, then we have $(N \cdot I) \rightarrow F$. By exportation, this is equivalent to $N \rightarrow (I \rightarrow F)$. In the context of N, we can query the conditional $I \rightarrow F$. And in N, we can *also* query its *consequent* F by taking I as a true presupposition when asking 'What happens in circumstances I?' N gives F as its answer to the *latter* question. Thus, N gives an affirmative answer to the question 'Does F happen in circumstances I?'

(*ii*) Since I *contradicts* E, we have both $(E \cdot I) \rightarrow F$ and $(E \cdot I) \rightarrow {\sim}F$. Equivalently, we have $E \rightarrow (I \rightarrow F)$ and $E \rightarrow (I \rightarrow {\sim}F)$. Thus, whereas E does give an affirmative answer when we query the conditional $I \rightarrow F$, it does *not* answer at all the question (2*b*) 'What happens under circumstances I?' or the question 'Does F happen under conditions I?', which queries the consequent F. For E entails *both* $I \rightarrow F$ and $I \rightarrow {\sim}F$ precisely because E declares I to be in principle unrealisable as a matter of physical law. In general, since $E \cdot I$ entails every *contrary* of F as well as F, E cannot answer the question (2*b*)! Thus, what sense would it make to say that E's obviation of this question is—in Popper's words—'an answer which is at least as precise' as N's answer F?

Hence the question (2*b*) is *not* answered in E but only *obviated as ill-conceived*. And this obviation obtains not despite but rather because of the fact that $I \in IC(E)$! It would be unavailing to baptise this insurmountable difficulty by saying that E 'answers' the question (2*b*) 'vacuously'. As well say that a man who never beat his wife at all can give a vacuous answer to the question as to whether and when he *stopped* beating her.

A *cadit quaestio* response is not an answer in the context of Popper's requirement of furnishing 'an answer which is at least as precise'.

Let us recall that according to Popper's proposed new *question-answering* criterion of comparative content for theories T_1 and T_2, T_1 has comparatively lower content than T_2 if both of the following conditions are satisfied [1972], pp. 52–3): (*a*) to every question to which T_1 has an answer, T_2 has an answer which is at least as precise, and (*b*) there are questions to which T_2 can give a (non-tautological) answer while T_1 does not. We shall refer to these two conditions as '*requirement* (*a*)' and '*requirement* (*b*)' respectively. Let us recall that we denote the (unmeasured) contents of T_1 and T_2 *in the sense of this new criterion, i.e.*, the respective sets of questions they answer, by '$QC(T_1)$' and '$QC(T_2)$' respectively and thereby distinguish them from the Tarskian contents $LC(T_1)$ and $LC(T_2)$ and the 'informative contents' $IC(T_1)$ and $IC(T_2)$ discussed earlier. When we are *not* concerned to distinguish between the (unmeasured) contents $QC(T)$, $LC(T)$ and $IC(T)$, we shall use the symbol '$Ct(T)$', where the *capital* '*C*' is used since Popper uses '$ct(T)$' with the lower case *c* to denote the content *measure* of T.

Now consider another dynamical example. In Einstein's theory of gravitation, the orbit of a planet of negligibly small mass *subject solely to the sun's gravitational field* is *not* a perfectly closed ellipse about the sun, as it is in Newton's theory. Instead, the orbit of such a planet is a slowly rotating ellipse. (For references, see my [1973], p. 577, n. 11). Now let us ask the question:

> (2*c*) Why is the orbit of a planet of negligible mass which is subject solely to the sun's field a perfectly closed ellipse about the sun?

Clearly this question *is* answered in N in the proper sense of exhibiting the presupposition of the question to be a consequence of N's theory of gravitation. In what sense, if any, does E 'answer' question (2*c*)? In the context of E, (2*c*) is predicated on a false presupposition and is thus an abortive non-starter. Can E therefore be *absolved* from having to *deal* with (2*c*) at all? And if so, can E be held to have 'answered' (2*c*) merely because E may be absolved from having to *deal* with (2*c*) at all?

It is true enough that it is *not* automatically incumbent upon a theory B to provide an answer to a question Q merely because Q is well-conceived and answered in a *rival theory* A in which the presupposition p of Q is held to be correct: If p is held to be either gratuitous or false in B but is asserted by A, then B's failure to answer Q in the sense of providing a *reason* for p is not necessarily a defect of B. By way of analogy, suppose that a court either has no information showing that a certain witness ever beat his wife at all or has information that the witness has never beaten his wife. Then it is true enough that the judge would hardly regard the witness as uncooperative when the witness fails to supply a date in answer to the question 'when did you last beat your wife?'

But how does this help Popper make good on the claim that question (2*c*) *is* 'answered' in E at least as precisely as it is in N, a claim which is essential to the fulfilment of requirement (*a*) of Popper's criterion? It would seem that E can be held to 'answer' (2*c*) only in the attenuated sense of declaring it to be ill-conceived in virtue of being predicated on a false presupposition: In Belnap's apt terminology, E '*obviates*' rather than *answers* (2*c*).

Suppose that Popper does countenance that such mere obviation by E of question (2*c*) counts as an 'answer' as he must if he is to disqualify (2*c*) from being a counter-example to the fulfilment of requirement (*a*) of his thesis $QC(N) \subset QC(E)$. How then will he prevent this defence from boomeranging by losing the latter inclusional asymmetry between the contents of N and E via having to allow N to merely obviate *at least any why-question* which E does answer in the stricter sense while N does not? Specifically, if mere obviation of a question also counts as 'answering' it, can Popper still demonstrate that E succeeds in fulfilling his requirement (*b*) *vis-à-vis* N, *viz.*, that there are questions to which E can give a (non-tautological) answer while N does not? The issue here is *not*, of course, whether there are questions to which E gives an *observationally correct* ('confirmed') answer while N does not; instead, it is whether there are questions to which N simply is logically unable to provide an answer—be it true or false—while E *can* answer, be it either truly or falsely. For the issue here is one of excess of *content*, not of 'observationally true' content! Thus, for example, it is immaterial here what the actual orbit of Mercury is when we compare the capabilities of N and E to answer why-questions as to the orbit of Mercury. In other words, it is not successfully tested content of a theory but only 'richness' of consequences that matter here.

Mindful of the latter, note for example, that if we ask the question

(2*d*) Why is the orbit of a planet of negligible mass which is subject solely to the sun's field a slowly rotating ellipse about the sun?

then N merely obviates (2*d*) while E answers it. But this fact does *not* enable Popper to adduce the case of (2*d*) as evidence *for* the fulfilment of the existential requirement (*b*), unless he *also* allows the following to count as evidence *against* the fulfilment of requirement (*a*): *E merely obviates* question (2*c*) while *N* does answer it! Furthermore, *N* and *E* can *each* answer any "yes-or-no" question of the form 'Is the orbit of Mercury such-and-such?' with great precision, and it is immaterial here that their incompatible answers may differ in truth value or approximate veridicality.

Furthermore, note that if *N* predicts a *zero* effect where *E* predicts a positive effect in a certain context (*e.g.*, gravitational shift of spectral lines), this fully allows for parity between them in regard to their ability to answer a question of the form 'Is there an effect of the specified kind in the amount *x*?' Only if *N* made *no* prediction at all—as distinct from a prediction of zero effect!—in a case where *E* does make a prediction could the case in point serve to show that the existential requirement (*b*) is being fulfilled by *E vis-à-vis N*. For in such a case *N* could *not* answer a yes-or-no question pertaining to the effect predicted by *E*. Nor would *N* then obviate either such a question or any why-question pertaining to this effect on the grounds that it is predicated on a false presupposition.

As it is, the very least we can conclude is that the burden of proof is on Popper to show that both the *universal* statement in requirement (*a*) and the existential statement in requirement (*b*) are fulfilled by *E* and *N*. Even if he can establish the fulfilment of (*b*) existentially, our doubts about (*a*) remain.

In our terminology, we can say that our illustrative sets of questions so far were offered as at least *prima facie* counterexamples to Popper's claim that $QC(N)$ and $QC(E)$ satisfy requirement (*a*) of his criterion for the proper inclusion of $QC(N)$ in $QC(E)$. It is important to note that (*i*) *E* gives *neither* the answer 'no' nor the answer 'yes' to *any* yes-or-no question which is ill-conceived in *E* though well-formed in *N and* predicated on assumptions which belong to $LC(N)$—as shown by the example of our question (1*a*), there are just such questions to which *N* does give a 'yes-or-no' type of answer, whereas *E* certainly does not—and (*ii*) in each of the two theories *N* and *E* respectively, there are well-formed 'why 'or 'how' questions *not* answered by the other and which do not call for a 'yes-or-no' type of answer.

Let me use the acronym '$PC(T)$' to denote Popper's 'problem content of a theory *T*' as defined by him in our quotation under 'Set 1'. And note that, as shown by the result (*ii*) just stated, the concept of $PC(T)$ is more restrictive than the concept of $QC(T)$, since $PC(T) \subset QC(T)$. Then it is clear from the result (*i*) just stated concerning question (1*a*) that $PC(N)$ and $PC(E)$ *fail* to satisfy Popper's requirement (*a*), *i.e.*, I have shown that $PC(N) \not\subset PC(E)$. Thus a construal of $Ct(T)$ as $PC(T)$ can no more vindicate Popper's claim that $Ct(N) \subset Ct(E)$ than a construal of $Ct(T)$ as $QC(T)$ can do so. Lest there be misunderstanding, let me emphasise that when commenting on his definition of $PC(T)$ by reference to $PC(N)$ and $PC(E)$, Popper himself ([1974], p. 21) does ***not*** claim that $PC(N) \subset PC(E)$ —he merely states that $y(N) \in PC(E)$ and $y(E) \in PC(N)$.

Set 3. So far we have confined ourselves to N and E. Let us now consider the comparison of the question–answering capabilities of the STR with those of another theory which incorporates Newtonian kinematics while going beyond it: the pre-STR Maxwell-Lorentz ether theory of light propagation, to be denoted by 'M-L'.

In the context of M-L, one can meaningfully ask and, in principle, answer questions such as

(3*a*) How is a light wave transmitted through the ether medium?

Here the STR denies the presupposition of the existence of the ether medium ***in the sense of the M-L theory***, and hence the STR does not answer ***this*** question, whatever other questions about light propagation it ***does*** answer.

Set 4. Let us now consider the feasibility of meeting requirement (*a*) of Popper's criterion for $QC(T_1) \subset QC(T_2)$ in the case of incompatible theories T_1 and T_2 which do not even differ ***conceptually*** in the way in which E differs from N, or in which the STR differs from M-L. A very interesting species of ***this kind*** of case is constituted by pairs of theories (T_1, T_2) such that T_2 is generated by modifying one of the auxiliary hypotheses in T_1 after a ***modus tollens*** argument has discredited T_1. It will be recalled from section **2** that our planetary theories τ_1 and τ_2 are an example of such a pair of theories: τ_1 asserted within the context of Newtonian mechanics that besides Uranus (and the Asteroids), there are ***only*** six planets all of which are ***intra***-Uranian, whereas τ_2 asserted that there exists an additional planet (Neptune) which is trans-Uranian.

As will likewise be recalled from **2**, Popper had claimed that $IC(\tau_1) \subset IC(\tau_2)$. And Popper invoked this purported content inclusion relation among these two theories as his reason for asserting the methodological legitimacy of the initial introduction of the auxiliary Neptune hypothesis on the basis of his general canon of content-increase. I have argued previously that the transition from τ_1 to τ_2 does NOT satisfy the sufficient condition of content-increase with respect to $IC(T)$, $EC(T)$ or $LC(T)$.

Hence it becomes quite important to ask whether Popper can now legitimate the introduction of the Neptune auxiliary hypothesis in accord with his general canon of content-increase by showing instead that either

P, 6584

$QC(\tau_1) \subset QC(\tau_2)$, or $PC(\tau_1) \subset PC(\tau_2)$, whence it would then follow that $Ct(\tau_1) \subset Ct(\tau_2)$ after all, as needed.

In the planetary theory τ_1, there is an (affirmative) answer to the question:

(4*a*) 'Does the distribution of the planets in regard to their distances from Uranus and the magnitudes of their masses account for the fact [*sic*!] that besides the sun *only* Jupiter and Saturn produce non-negligible gravitational effects on the observable motion of Uranus?'

But in the modified planetary theory τ_2, there can be no answer to *this* question: In the latter theory, (4*a*) is an abortive non-starter. For precisely by postulating Neptune, τ_2 denies the purported 'fact' on which (4*a*) is predicated. And if Popper is allowed to plead mere obviation of (4*a*) by τ_2 in the interest of fulfilling requirement (*a*), how then can Popper demonstrate the fulfilment of the existential requirement (*b*) amid allowing a like obviation of questions by τ_1?

As for other yes-or-no questions, how is he going to demonstrate that $PC(\tau_1) \subset PC(\tau_2)$? And even if he can do so in *this* case, how can he do so *generally* in such legitimate cases of theory-modification?

4 CONCLUSION

The discussion I have given supports the contention that one or more of the following criticisms of Popper's Thesis Q are justified:

(1) In the case of empirically competing (incompatible) and yet conceptually disparate theories having different ontologies,

(*i*) Any question asked in either theory which is predicated on an assumption denied by the other or not even *statable* in the other's conceptual framework will not be answered by the other (with at least equal precision) but at best will be obviated by the other,

(*ii*) Contrary to Popper, it is not the respects in which N and E converge *qua* N's being a limiting case of E, but rather the respects in which they *diverge*, that are decisive in regard to whether $QC(N)$ is properly included in $QC(E)$,

(*iii*) Just as we found the contents of *incompatible* theories to be incomparable with respect to proper inclusion when $Ct(T)$ was construed as $IC(T)$ or $LC(T)$, so also there is such failure of proper inclusion if $Ct(T)$ is construed as $QC(T)$ or $PC(T)$ in the cardinally important case which Popper had adduced to the contrary.

(*iv*) Glymour's global topology example would seem to show that countenancing obviation of a question as being an 'answer' does *not* disarm *this* example as a counterexample to the fulfilment of Popper's requirement (*a*) by E.

(2) In the case of our two planetary theories τ_1 and τ_2 (or of the *M-L* ether theory and the Lorentz–Fitzgerald ether theory), the claim that $PC(\tau_1) \subset PC(\tau_2)$ would appear to be gratuitous if not false. And thus even when we construe content as problem-content, Popper has supplied no basis for *justifying* the introduction of modifying auxiliary hypotheses in the typical or important historical cases on the basis of his canon of content-increase. Also, it is at best moot whether he can certify compliance with his criterion of admissibility by means of his *dimension* method for comparing degrees of falsifiability. Thus, it is at best unclear whether Popper's criterion of admissibility for auxiliaries has any significant applications at all to concrete cases.

University of Pittsburgh

REFERENCES

Grünbaum, A. [1973]: *Philosophical Problems of Space and Time*, Second enlarged edition.

Grünbaum, A. [1976]: 'Is falsifiability the touchstone of scientific rationality? Karl Popper versus inductivism', in R. S. Cohen and M. Wartofsky (*eds.*): *Lakatos Memorial Volume, Boston Studies in the Philosophy of Science*, 39.

Havas, P. [1964]: 'Four-dimensional formulations of Newtonian mechanics and their relation to the special and the general theory of relativity', *Reviews of Modern Physics*, 36, pp. 938–65.

Miller, D. [1974a]: 'Popper's qualitative theory of verisimilitude', *Journal of Symbolic Logic*, 39, pp. 414–15.

Miller, D. [1974b]: 'Popper's qualitative theory of verisimilitude', *The British Journal for the Philosophy of Science*, 25, pp. 166–77.

Miller, D. [1975]: 'The Accuracy of Predictions', *Synthèse*, 30, 1/2, 1975, pp. 159–91.

Newton, I. [1934]: *Principia Mathematica, ed.* by F. Cajori.

Popper, K. R. [1959]: *The Logic of Scientific Discovery.*

Popper, K. R. [1962]: *Conjectures and Refutations.*

Popper, K. R. [1966]: 'A theorem on truth-content', in P. K. Feyerabend and G. Maxwell (*eds.*): *Mind, Matter, and Method*, pp. 343–53.

Popper, K. R. [1972]: *Objective Knowledge.*

Popper, K. R. [1974]: 'Intellectual Autobiography' and 'Replies to my critics' in P. A. Schilpp (*ed.*): *The Philosophy of Karl Popper*, Library of Living Philosophers.

Schrödinger, E. [1956]: *Expanding Universes.*

Tichý, P. [1974]: 'On Popper's definitions of verisimilitude', *The British Journal for the Philosophy of Science*, 25, pp. 155–88.

Brit. J. Phil. Sci. 27 (1976), 25–42 *Printed in Great Britain*

Verisimilitude Redefined

by PAVEL TICHÝ

1 *Examples.*
2 *Intuitive Background.*
3 *The Definition.*
4 *The Examples Revisited.*
5 *A Look Back on Popper's Approach.*

Are some falsehoods less false than others? Tarski's definition of truth tells us in rigorous terms what it takes for a statement to be *false*. Can we have an equally rigorous definition of what it takes for one false statement to be *closer to the truth* than another?

The notion of verisimilitude is bound to figure centrally in any adequate theory of scientific progress. For in the empirical sciences change invariably takes the form of one false theory being discarded in favour of another false theory. If such a development is to be meaningfully qualifiable as a step forward, it must make sense to say that the new theory, although also false, is closer to the truth than its discarded predecessor.

The problem of verisimilitude was first explicitly raised by Sir Karl Popper. Popper has repeatedly emphasised the fundamental importance of the notion and the urgent need for its rigorous explication.[1] However, Popper's own attempts at giving an explication have failed.[2] The aim of the present paper is to propose a viable definition of the concept.

1 EXAMPLES

The intuition underlying the concept of verisimilitude is a remarkably distinct one. In order to elicit this intuition in the reader, I shall cite three very simple examples.

(Two notational conventions: a homogenous string of quantifiers, say '$(\forall x_1) \ldots (\forall x_n)$' will be abbreviated to '$(\forall x_1 \ldots x_n)$'; a dot in a formula will represent a left-hand bracket whose right-hand mate is to be imagined as far to the right as is consistent with other pairs of brackets.)

Received 5 *August* 1975

---

[1] See Popper [1962], Popper [1963], and Popper [1972].

[2] Popper has proposed two definitions of verisimilitude: one in terms of logical consequence and one in terms of logical probability. The former has been found vacuous independently by David Miller and myself (see Miller [1974] and Tichý [1974]). The latter has been shown inadequate in Tichý [1974].

EXAMPLE A. Consider a language L_A whose only extralogical primitive symbols are three independent sentences h, r, and w. Assume that all the three sentences are, as a matter of fact, true. The following are then four false theories in L_A:

$(A1)$ $\quad \sim h \,\&\, \sim r \,\&\, \sim w$
$(A2)$ $\quad \sim h \,\&\, r \,\&\, \sim w$
$(A3)$ $\quad \sim h \,\&\, r$
$(A4)$ $\quad \sim h \,\&\, r \,\&\, w$

Theory $(A1)$ gets everything wrong; its is the worst consistent theory one can propose in L_A. $(A2)$ is also rather bad, but definitely better than $(A1)$: it gets at least something right. $(A3)$ shares all $(A2)$'s merits, but only some of its deficiencies: in one point where $(A2)$ is wrong, $(A3)$ suspends judgment, thus allowing for the right answer. It is thus plausible to say that $(A3)$ is marginally better than $(A2)$. $(A4)$ is almost right; it clearly surpasses all the other theories. In a word, $(A1)$–$(A4)$ are (in this order) increasingly better approximations to the truth.

EXAMPLE B. Let L_B be a first-order language (with identity) whose alphabet α_B of extralogical primitive symbols consists of two independent one-place predicates P and Q, and whose universe of discourse is a denumerably infinite collection ι_B. Assume that in actual fact P and Q have two-element extensions which partially but not entirely overlap (thus having exactly one element in common). We shall write '$(\exists_n x)A$' for the standard formula saying that exactly n values of x satisfy A. Thus for instance, '$(\exists_2 x)Px$' is short for '$(\exists xy)\,.\sim x = y \,\&\, Px \,\&\, Py \,\&\, (\forall z)\,.\,Pz \supset .\, z = x \vee z = y$'. The following are four false theories in L_B:

$(B1)$ $\quad (\exists_1 x)Px \,\&\, (\forall x) \sim Qx$
$(B2)$ $\quad (\exists_2 x)Px \,\&\, (\forall x) \sim Qx$
$(B3)$ $\quad (\exists_2 x)Px \,\&\, (\exists_1 x)Qx \,\&\, (\forall x) \supset \,\sim Qx$
$(B4)$ $\quad (\exists_2 x)Px \,\&\, (\exists_2 x)Qx \,\&\, (\forall x)\,.\,Px \supset \,\sim Qx.$

Theory $(B1)$ is very bad. It ascribes P a one-element extension and Q an empty one. $(B2)$ corrects the former error of $(B1)$ but shares the latter. $(B3)$ partakes in all the assets of $(B2)$ and mitigates its errors: while $(B2)$ made Q empty, $(B3)$ sets the number of Q's elements at 1. Thus $(B1)$–$(B3)$ are increasingly better approximations to the truth. As for $(B4)$, it corrects something which $(B3)$ gets wrong, *viz.* the size of Q; on the other hand, it gets wrong something which $(B3)$ gets right, *viz.* the number of Q's which are outside P. $(B3)$ and $(B4)$ thus seem equidistant from the truth.

EXAMPLE C. Finally, consider a language L_C whose alphabet α_C consists of one two-place predicate R and two individual constants o and n, and whose universe ι_C is again a denumerably infinite set. Assume that, as a matter of fact, the extension of R effects an irreflexive linear ordering of ω and that o and n are, respectively, the first and the fifth term of that ordering. Consider the formula

(γ) $(\forall xyz) . [\sim x = y \supset . Rxy \vee Ryx] \,\&\, [Rxy \,\&\, Ryz \supset Rxz] \,\&\, [Rxy \supset \sim Ryx] \,\&\, Rox.$

In order to save space, let us write '$R^*(x,y)$' for '$Rxy \,\&\, (\forall z) \sim . Rxz \,\&\, Rzy$' The following are four false theories in L_ω:

(C1) $\gamma \,\&\, o = n$
(C2) $\gamma \,\&\, R^*(o,n)$
(C3) $\gamma \,\&\, (\exists x) . R^*(o,x) \,\&\, R^*(x,n)$
(C4) $\gamma \,\&\, (\exists xy) . R^*(o,x) \,\&\, R^*(x,y) \,\&\, R^*(y,n)$.

Again, (C1)–(C4) are increasingly better approximations to the truth. Each of the theories is right in affirming that R is a linear ordering and that o comes first in that ordering. Now the remaining, false, information they give is that n is, respectively, the first, second, third, and fourth term of the ordering. But to say, successively, of the fifth member of an ordering that it is the first, second, third, and fourth member, is clearly zeroing in on the truth.

Consider, moreover, the following three theories:

(C5) (C1) $\vee$ (C2) $\vee$ (C3) $\vee$ (C4)
(C6) (C2) $\vee$ (C3) $\vee$ (C4)
(C7) (C3) $\vee$ (C4).

Theory (C5) places n between the first and fourth member of the ordering (including the extremes), (C6) places it between the second and the fourth and (C7) between the third and the fourth. (C5)–(C6) are thus markedly less specific than (C1)–(C4), and the differences among them as to proximity to the truth are, accordingly, somewhat less striking. My intuition tells me nevertheless that (C6) is closer to the truth than (C5) and that (C7) is closer to the truth than (C6). My intuition is here in full agreement with Popper's. Popper has considered[1] the statements 'It is now between 9.45 and 9.48' and 'It is now between 9.40 and 9.48' where 'between' is supposed to exclude the two bounds. Supposing that the actual time is 9.48, Popper says, 'we can . . . assert that the first [statement] has a greater truthlikeness than the second'. It is readily seen that the way Popper's two statements are related to the truth is completely analogous to the way (C7) and (C6) or (C7) and (C5) are related to the truth in our example (n corresponding, of course, to Popper's 'now').

How do (C5)–(C7) compare with (C1)–(C4)? Both (C5) and (C6) seem definitely better than ($C1$) and ($C2$). Both allow for the very bad estimates of *n*'s position made by ($C1$) and ($C2$), but all the other estimates they also allow for are significantly better. ($C6$) locates *n* roughly where ($C3$) does; but being, as it is, substantially vaguer about it than ($C3$), it seems marginally worse than ($C3$). But ($C7$) is definitely better than ($C3$); it allows for ($C3$)'s quite close guess, and, besides, for a better one. ($C4$), on the other hand, is patently better than ($C7$). Intuitively, therefore, the ordrer of increasing excellence is this: ($C1$), ($C2$), ($C5$), ($C6$), ($C3$), ($C7$), ($C4$).

---

[1] See Popper [1972], p. 56.

2 INTUITIVE BACKGROUND

The problem of explicating verisimilitude is essentially a normalisation problem. Each first-order formula, *i.e.* each first-order theory, is a finite object, and the amount of information it carries is accordingly finite. It seems intuitively clear that finite pieces of information can always be judged as to how 'close' or 'similar' they are to one another. If this intuitive notion is correct, no insurmountable difficulties should stand in the way of a rigorous definition of truthlikeness: one false theory can be naturally defined to have a greater truthlikeness than another if it surpasses this other theory in closeness to the theory which happens to be true. All that is needed is a method of codifying the information carried by first-order formulas into such a form as will make such comparisons possible. A natural method of this sort, however, can be easily extracted from Jaakko Hintikka's theory of distributive normal forms.[1]

Consider a first-order language L having a finite vocabulary α of predicates and individual constants, and a fixed universe of discourse ι. Where A is a formula of L, let us ask: exactly what is it that A says? Exactly what is the (correct or incorrect) information it conveys?

One, trivial, answer to this question is, of course, in terms of the familiar Tarskian truth-definition. What one does is simply to *translate* A as it stands into the metalanguage. Where quantifiers occur in A, the metalinguistic translation will involve quantification over the whole universe ι. This is why Hintikka speaks of this approach to answering the above question as a *global* one.

For many purposes, however, it is much more rewarding to take what Hintikka aptly terms a *local* point of view. On the local approach what one asks is roughly this: what sort of actual experiences could one expect if one set out to explore a world in which A is true?

Now actual experiences are always concerned with not more than a finite number—say d—of individuals; one cannot survey, as it were, the whole universe ι at a glance. Each particular experience can thus be imagined as an inspection of a string of d individuals, or briefly a *d-string*. Such an inspection reveals what properties are instantiated by the components of the d-string and what relations hold among the components. In view of the fact that the vocabulary α of L is finite, the total amount of information resulting from the inspection is finite. In other words, the inspection results in a classification of the string into one of a finite number of possible (and mutually exclusive) kinds of d-string, or briefly *d-kinds*. The question one asks on the local approach is then more particularly this: what d-kinds can one expect to come across in a world where A is true?

---

[1] See Hintikka [1963].

Now let $\phi = \{K_1, \ldots, K_m\}$ be an arbitrary family of d-kinds, or briefly a *d-family*. There may or may not be possible worlds—*i.e.* interpretations of the symbols from α in ι compatible with the meaning of those symbols—which feature d-strings of the kinds $K_1, \ldots, K_m$ and of no other kinds. If there are such worlds, let us call them *ϕ-worlds* and ϕ itself a *consistent* d-family. It is readily seen that there is a finite set ψ of consistent d-families such that for each possible world W there is one and only one ϕ in ψ such that W is a ϕ-world. In particular, there is a unique d-family, call it ϕ^*, such that the actual world is a ϕ^*-world.

A formula is said to be of *depth d* if the longest sequence of nested quantifiers (*i.e.* a sequence whose every member stands within the scope of each of its predecessors) ocurring in the formula is of length d. Now it is not difficult to see that for any consistent d-family $\phi = \{K_1, \ldots, K_m\}$ there is in L a formula C^ϕ of depth d which says no more and no less than that the actual world is a ϕ-world, *i.e.* that $K_1, \ldots, K_m$ is a full list of exemplified d-kinds. Formulas of this sort are known as *d-constituents*.

Now it follows from Hintikka's theorem on distributive normal forms that for any consistent formula A of depth d there are consistent d-families $\phi_1, \ldots, \phi_s$ such that A is equivalent to $C^{\phi_1} \vee \ldots \vee C^{\phi_s}$. The assertive content of A is thus seen to be to the effect that the (actual) world is a ϕ_1-world, or . . . or a ϕ_s-world; the set $\{\phi_1, \ldots, \phi_s\}$ can thus serve as a normalisation or codification of the informative content carried by A. And it is this codification which reveals similarities and dissimilarities among such contents.

If we had a plausible way of measuring distances between d-families, the distance of A from the truth would be naturally defined as the average distance of the members of the corresponding set $\{\phi_1, \ldots, \phi_s\}$ from the 'actual' d-family ϕ^*. Let us see therefore, if we can find an acceptable way of determining distances between d-families. It will be recalled that *d*-families are sets of *d*-kinds. Now imagine that we know how to measure distances between *d*-kinds. A natural way of determining the distance from *d*-family ϕ' to *d*-family ϕ would be by finding the most economical way of pairing off the members of ϕ with members of ϕ' and taking the average distance between the members of the pairs. Our problem is thus reduced to that of defining, in general, the distance between two *d*-kinds.

A *d*-kind, we have seen, is a complete description, in terms of α, of a string of *d* individuals. It can thus be regarded as a sequence of answers to a finite number of yes/no questions, such as 'Does the first component of the string instantiate such-and-such property?', 'Does the first component bear such and such relation to the second component?' and so on.

But a natural way of measuring similarities and dissimilarities among such sequences suggests itself: two such sequences are the more similar or 'close' the more concurrent answers they give. The 'distance' between the two sequences can be naturally identified with the number of divergent answers they give.

In the next section these informal considerations will be translated into rigorous terms.

3 THE DEFINITION

1. *Definition.* Let $K = \langle k_1, \ldots, k_n \rangle$ and $K' = \langle k'_1, \ldots, k'_n \rangle$ be two ordered n-tuples of zeroes and ones. The *distance* $\epsilon(K',K)$ *of* K' *from* K is the cardinality of $\{i \mid 1 \leqslant i \leqslant n \text{ and } k'_i \neq k_i\}$.

2. *Definition.* Let R be a binary relation between ordered n-tuples of zeroes and ones. The ***breadth*** $\beta(R)$ *of* R is the average distance of the second relatum of R from the corresponding first relatum. Thus

$$\beta(R) = \Sigma\{\epsilon(K',K) \mid RKK'\}/\overline{\overline{R}}$$

3. *Definition.* Let ϕ and ϕ' be two non-empty sets of n-tuples of zeroes and ones. Relation R is said to be a *linkage from* ϕ' *to* ϕ if ϕ' is the domain and ϕ the range of R, and no relation of cardinality smaller than that of R has the same domain and range as R has.

4. *Definition.* Where ϕ and ϕ' are as in 3, the *distance* $\delta(\phi',\phi)$ *of* ϕ' *from* ϕ is the breadth of the narrowest linkage from ϕ' to ϕ. Thus $\delta(\phi',\phi) = \min b$ [there is a linkage R from ϕ' to ϕ such that $b = \beta(R)$].

5. *Definition.* Let Φ be a class of classes of n-tuples of zeroes and ones and let ϕ be as in 3. The *distance* $\delta'(\Phi,\phi)$ *of* Φ *from* ϕ is the average distance of the members of Φ from ϕ. Thus

$$\delta'(\Phi,\phi) = \sum\{\delta(\phi',\phi) \mid \phi' \in \Phi\}/\overline{\overline{\Phi}}.$$

We shall now consider a first-order language L (with identity). Let α be the extra-logical *vocabulary* of L (consisting of a finite number of propositional, predicate, and individual constants), ι the *universe of discourse* and ω the *logical space* of L. Thus ω is the set of *possible worlds*, *i.e.* of interpretations of α in ι compatible with the meaning of the symbols in α. Assume that W^* is the actual world.

Let $x_1, \ldots, x_d$ be the first d (individual) variables of L.

6. *Definition.* By $I(x_1, \ldots, x_d)$ we shall understand the conjunction of all the formulas of the form $\sim y = x$, where x is one of $x_1, \ldots, x_d$ and y is an individual constant in α or one of $x_1, \ldots, x_d$ other than x. Let us stipulate that of any two conjuncts of the form $\sim x = y$ and $\sim y = x$ one has been deleted from $I(x_1, \ldots, x_d)$.

7. *Definition.* Let $B_1(x_1, \ldots, x_d)$, $B_2(x_1, \ldots, x_d), \ldots, B_n(x_1, \ldots, x_d)$ be a full list (lexicographically ordered) of all atomic formulas constructible from α and $x_1, \ldots, x_d$, except for those occurring in $I(x_1, \ldots, x_d)$ and those of the form $x = x$; let us also stipulate that of two formulas of the form $x = y$ and $y = x$ only one appears in the list.

8. *Definition.* Any ordered n-tuple of zeroes and ones is a *d-kind.*

9. *Definition.* Where $K = \langle k_1, \ldots, k_n \rangle$ is a d-kind, by $M^K(x_1, \ldots, x_d)$ we shall understand an n-way conjunction whose j-th conjunct ($1 \leqslant j \leqslant n$) is $B_j(x_1, \ldots, x_d)$ or $\sim B_j(x, \ldots, x_d)$ according as k_j is zero or one.

10. *Definition.* Where K is as in 9, by $IM^K(x_1, \ldots, x_d)$ we shall understand $I(x_1, \ldots, x_d)$ & $M^K(x_1, \ldots, x_d)$.

EXAMPLE B (continued). The following is a lexicographically ordered list of atomic formulas constructible from α_B and the variables x_1, x_2, x_3:

$$Px_1,\quad Px_2,\quad Px_3,\quad Qx_1,\quad Qx_2,\quad Qx_3.$$

Thus, relative to L_B, 3-kinds are ordered sextuples of zeroes and ones. Let K be, for instance, the 3-kind $\langle 0,1,1,0,0,1 \rangle$. Then $IM^K(x_1,x_2.x_3)$ is the formula

$$\sim x_1 = x_2 \;\&\; \sim x_1 = x_3 \;\&\; \sim x_2 = x_3 \;\&\; Px_1 \;\&\; \sim Px_2 \;\&\; \sim Px_3 \;\&\; Qx_1 \;\&\; Qx_2 \;\&\; \sim Qx_3.$$

11. *Definition.* Classes of d-kinds are called *d-families.*

12. *Definition.* Let $\phi = \{K_1, \ldots, K_m\}$ be a d-family. A possible world W in ω is said to be a *ϕ-world* if for any d-kind K,

$$K \in \phi \text{ iff } IM^K(x_1, \ldots, x_d) \text{ is satisfiable in } W.$$

EXAMPLE B (continued). Consider an interpretation W_1 of α_B in ι_B which assigns a one-element set to P and the empty set to Q. W_1 is clearly a ϕ_1-world, where ϕ_1 has the following four elements:

$K_1^3 = \langle 0,1,1,1,1,1 \rangle$, $K_2^3 = \langle 1,0,1,1,1,1 \rangle$, $K_3^3 = \langle 1,1,0,1,1,1 \rangle$, and
$K_4^3 = \langle 1,1,1,1,1,1 \rangle$.

13. *Definition.* A d-family ϕ is said to be ***consistent*** if ω contains a ϕ-world.

14. *Definition.* Let ϕ be as in 12. By C^{ϕ} we shall understand the following formula:

$$(\exists x_1 \ldots x_d)IM^{K_1}(x_1, \ldots, x_d) \;\&\; (\exists x_1 \ldots x_d)IM^{K_2}(x_1, \ldots, x_d) \;\&\; \ldots$$
$$\&\; (\exists x_1 \ldots x_d)IM^{K_m}(x_1, \ldots, x_d) \;\&\; (\forall x_1 \ldots x_d) \,.\, I(x_1, \ldots, x_d) \supset$$
$$.M^{K_1}(x_1, \ldots, x_d) \vee M^{K_2}(x_1, \ldots, x_d) \vee \ldots \vee M^{K_m}(x_1, \ldots, x_d).$$

15. *Proposition.* C^{ϕ} is true in world W iff W is a ϕ-world.

The following proposition is an easy consequence of 15 and Hintikka's existence theorem for distributive normal forms in languages with identity.

16. *Proposition.* Let $d \leqslant \bar{\imath}$ and let A be a closed formula of depth less than or equal to d, true in at least one world in ω. Then there is a unique set $\Phi = \{\phi_1, \ldots, \phi_s\}$ of consistent d-families such that A is equivalent to $C^{\phi_1} \vee \ldots \vee C^{\phi_s}$.

17. *Definition.* Let d, A, and Φ be as in 16. Then Φ is called the *d-content* of A.

EXAMPLE B (continued). It is easy to check that theory (B1) is equivalent to C^{ϕ_1}, where ϕ_1 is the 3-family specified above. Thus the 3-content of (B1) is $\{\phi_1\}$.

18. *Definition.* Let d, A, and Φ be as in 16. The *d-distance* $\triangle_d(A,W)$ *of A from* world W is $\delta'(\Phi,\phi)$, where ϕ is the d-family such that W is a ϕ-world.

EXAMPLE B (continued). Let us determine the 3-distance of formula (B2)$\vee$(B3) from the world W_1 specified above. It is easy to check that (B2) and (B3) are equivalent, respectively, to the constituents C^{ϕ_2} and C^{ϕ_3}, where ϕ_2 has 7 members, *viz.* K_1^3—K_4^3 (see *above*) $K_5^3 = \langle 0,0,1,1,1,1 \rangle$, $K_6^3 = \langle 0,1,0,1,1,1 \rangle$, and $K_7^3 = \langle 1,0,0,1,1,1 \rangle$, and ϕ_3 has 19 members, *viz.* K_1^3—K_7^3, $K_8^3 = \langle 0,0,1,1,1,0 \rangle$, $K_9^3 = \langle 0,1,0,1,0,1 \rangle$, $K_{10}^3 = \langle 0,1,1,1,0,1 \rangle$, $K_{11}^3 = \langle 0,1,1,1,1,0 \rangle$, $K_{12}^3 = \langle 1,0,0,0,1,1 \rangle$, $K_{18}^3 = \langle 1,0,1,0,1,1 \rangle$, $K_{14}^3 = \langle 1,1,0,0,1,1 \rangle$, $K_{15}^3 = \langle 1,1,1,0,1,1 \rangle$, $K_{16}^3 = \langle 1,0,1,1,1,0 \rangle$, $K_{17} = \langle 1,1,0,1,0,1 \rangle$ $K_{18}^3 = \langle 1,1,1,1,0,1 \rangle$, and $K_{19}^3 = \langle 1,1,1,1,1,0 \rangle$. We know that W_1 is a ϕ_1-world. Thus to obtain the distance of ϕ_2 from ϕ_1 we must first find an optimal linkage from ϕ_2 to ϕ_1. One such linkage is clearly $\{\langle K_1^3,K_1^3 \rangle$, $\langle K_2^3,K_2^3 \rangle$, $\langle K_3^3,K_3^3 \rangle$, $\langle K_4^3,K_4^3 \rangle$, $\langle K_5^3,K_2^3 \rangle$, $\langle K_6^3,K_1^3 \rangle$ $\langle K_7^3,K_2^3 \rangle\}$. The distances between the components of the ordered couples forming the linkage are, respectively, 0, 0, 0, 0, 1, 1, 1. The breadth of the linkage is therefore 3/7. Thus $\delta(\phi_2,\phi_1) \simeq 0.43$. In a similar fashion it can be determined that $\delta(\phi_3,\phi_1) \simeq 0.84$. Consequently, $\triangle_3((B2)\vee(B3),W_1) = \delta'(\{\phi_2,\phi_3\},\phi_1) \simeq 0.63$.

The procedure just exemplified was briefly sketched in Tichý [1974].

David Miller,[1] commenting upon that sketch, speaks disparagingly of 'Tichý's "counting method"'. The implication seems to be that it is not proper for a logician to stoop to counting things.

It is, however, one of the facts of life that one either counts, or one does not arrive at a *number*. Thus if one wants to arrive at a numerical measure of verisimilitude, one has to count something or other. Popper's way of computing his measure of verisimilitude is no less a counting method than mine. His definition is in terms of logical probability and, as well known, the logical probability of a proposition is arrived at by *counting* the possible worlds in which the proposition holds true.

Popper's probabilistic definition of verisimilitude is inadequate not because it involves counting, but because it directs us to count things of the wrong sort. Verisimilitude cannot be defined in terms of the number of *worlds* in which the formulas in question come true. All this number tells us is how *strong* the formulas are; and it is evident that of two equally strong formulas, one can be almost right and the other disastrously wrong. In the logical probability of a formula, all that makes the formula close or distant from the truth is lost.

If one wants to obtain a measure of verisimilitude, one must not count vertically across the logical space. Counting must rather proceed horizontally, *within* individual possible worlds. One must attend to atomic states of affairs constituting the worlds in which the formula holds and contrast them with atomic states of affairs which actually obtain.

The explication of the intuitive notion of verisimilitude I wish to propose is now as follows:

19. *Definition.* Let A and A' be arbitrary formulas of L and let $d \leqslant \bar{1}$ be the depth of A or the depth of A', whichever is greater. *A' has more verisimilitude than A* iff $\triangle_d(A',W^*) < \triangle_d(A,W^*)$.

4 THE EXAMPLES REVISITED

EXAMPLE A (continued). On Definition 18, the distances of ($A1$), ($A2$), ($A3$), and ($A4$) from the actual world are 3, 2, 1.5, and 1 respectively; thus on 19, they have increasingly more verisimilitude, as intuitively expected.

It seems to me that if ($A1$) and ($A4$) do not constitute an instance of one theory being closer to the truth than another, then nothing does. In Tichý [1974], therefore, I took it for granted that in order to demonstrate that Popper's probabilistic definition of verisimilitude is inadequate, it is sufficient to show that the definition makes ($A1$) and ($A4$) equidistant from the truth. Yet David Miller, commenting on my paper in Miller [1974], saw fit to take me to task for this and to insist that ($A4$) is indeed no closer to the truth than ($A1$). He propped up his startling claim with a somewhat intractable semantic argument (see below) which, if generalised, purports to show that it is impossible to 'make . . . correct judgments of . . . comparative proximity to the truth' at all. In particular, all the intuitive judgments of comparative verisimilitude we have made in Examples *A*, *B*, and *C* are, as far as Miller is concerned, completely spurious.

---

[1] Miller [1974], p. 176.

One way to construe Miller's position is as a confession that he, Miller, simply lacks my intuition that ($A4$) is closer to the truth than ($A1$), that he lacks Popper's and my intuition that ($C7$) is closer to the truth than ($C5$), and so forth. If this construal was correct, what Miller had to say would amount to no more than a piece of his own autobiography, regrettable, but of little philosophical interest. What more, it would be quite inexplicable why he should consider verisimilitude a worthwhile topic in the first place: trying to explicate a putative notion which lacks a single clearcut intuitive instance, is hardly a serious venture. Finally, it would be totally incomprehensible why Miller should find Popper's 'qualitative' definition of verisimilitude 'quite untenable'.[1] For implying, as it does, that no false theory is closer to the truth than another theory, the definition would be rather in perfect agreement with what Miller intuitively expected.

But if, on the other hand, Miller shares Popper's and my pretheoretical intuitions about verisimilitude—intuitions, by the way, which I have so far found in the intellectual armoury of every person I have spoken to—then he has misunderstood the import of his own semantic argument (mentioned above and quoted below). What that argument demonstrates is that on the semantic precepts which underlie the argument it is impossible to do justice to the pretheoretical notion of truthlikeness we all have. The obvious conclusion seems to be that some of those precepts need revising. But Miller concludes instead that the intuitive notion of truthlikeness should be given up. Thus an ichthyologist might urge us to give up the idea that fish can swim, pointing out that his theory of fish behaviour prevents it.

It is not very difficult to see which of Miller's semantic presuppositions is in need of revision.

In Tichý [1974] I imagined h, r, and w to express, respectively, the propositions 'It is hot', 'It is raining', and 'It is windy'. I envisaged L_A being used by 'Jones and Smith, two prisoners sharing a windowless and air-conditioned cell, . . . to discuss the weather'. In particular, Jones was assumed to conjecture ($A1$) and Smith ($A4$). Miller's comments are now as follows:

---

[1] Miller [1974], p. 166.

> ... is Tichý right to suppose it 'hardly deniable' that Smith is nearer to the truth than Jones is? I do not really think so. Consider another rudimentary weather language, trivially intertranslatable with [L_A]. This language also has three atomic sentences. One is *h* ('It is hot'). Another is *m* ('It is Minnesotan'); the third is *a* ('It is Arizonan'). The weather in Minnesota, we may suppose, is either hot and wet or cold and dry. Thus *m* is equivalent to $h \Leftrightarrow r$. Arizonan weather, on the other hand, is either hot and windy or cold and still. Thus *a* is equivalent to $h \Leftrightarrow w$. The three sentences *h*, *m*, *a* are logically independent of one another; and the eight constituents that can be formed from the three generators *h*, *m*, *a* are just the eight constituents that can be formed from *h*, *r*, *w*. Moreover, *r* is equivalent to $h \Leftrightarrow m$, and *w* is equivalent to $h \Leftrightarrow a$. There is perfect symmetry. And there seems no good reason—beyond sheer prejudice—for treating the *h–r–w* language as more fundamental than the *h–m–a* one. Yet Smith's theory in the latter language is $\sim h \;\&\; \sim m \;\&\; \sim a$, as is easily checked, whilst Jones's is $\sim h \;\&\; m \;\&\; a$. The truth comes out as $h \;\&\; m \;\&\; a$. Thus Smith is wrong on three counts. Jones on only one. Which one now is nearer to the truth?
>
> What does seem to me 'hardly deniable' is that the theories $h \;\&\; m \;\&\; a$ and $h \;\&\; r \;\&\; w$ are equivalent theories.

I, for one, find this last contention quite easy to deny. Two statements are equivalent if they say the same thing, *i.e.* if they have the same affirmative force. Now what is the affirmative force of a statement? It has been a commonplace at least from Carnap's *Meaning and Necessity* to identify this force with the *range* of that statement, *i.e.* with the class of possible states of affairs in which the statement holds true. Two statements are equivalent just in case they have the same range. What is a possible state of affairs? It is a distribution of truth-values through the class of atomic propositions of the language in hand, a function, that is, which takes the atomic propositions to the truth-values. The totality of such functions is known as the logical space of the language. Thus if the ranges of two statements are to be identical, their elements must come from the same logical space. Now $h \;\&\; r \;\&\; w$ is a statement in *h–r–w*–ese, hence its range consists of functions from $\{h,r,w\}$ to the truth-values; $h \;\&\; m \;\&\; a$, on the other hand, is in *h–m–a*–ese, hence its range consists of functions from $\{h,m,a\}$ to the truth values. Since $\{h,r,w\}$ and $\{h,m,a\}$ are clearly two distinct sets of propositions, no function defined on the former set can be identical with one defined on the latter. Thus the ranges of $h \;\&\; r \;\&\; w$ and $h \;\&\; m \;\&\; a$, far from being identical, have in fact nothing in common. Miller may well be right in maintaining that it would be sheer prejudice to treat *h–r–w*–ese as more fundamental than *h–m–a*–ese. But Miller's own view that *h–r–w*–ese is *equivalent* to *h–m–a*–ese definitely stems from an untenable prejudice. The two languages have completely different logical spaces and therefore no statement made in one of them is translatable into the other.

The prejudice which prompts Miller to deny this simple fact is to the effect that logical space is not to be taken seriously. To take it seriously would be, on Miller's view, making verisimilitude and truth itself intolerably language-dependent. In Miller's own words,

> Verisimilitude, like truth, will always, if defined, be defined relative to a language. But just as truth is language-independent . . . so must judgments of verisimilitude be, if they are to have any objective significance at all.

—a juxtaposition of statements I fail to understand. Truth and verisimilitude are, of course, language-independent in that they cannot be excogitated from the mere *definition* of the language in hand; extra-linguistic matters decide which (non-trivial) statements are true and which are false, and which false statements are close to the truth and which are far from it. But the assertive *force* of those statements is an exclusively linguistic matter, something determined entirely by the definition of the language. And it would be absurd to demand that the truth and verisimilitude of a statement should be independent of its assertive force, *i.e.* of what the statement *says*.

Miller's argument and his further researches reported in his as yet unpublished article 'The Accuracy of Predictions'* are nevertheless of considerable interest. What they show is that we face a choice: we either do take logical space seriously or we may as well give up hope of ever making sense of our ingrained intuitive notion of truthlikeness. This dilemma, however, does not strike me as particularly difficult to resolve.

EXAMPLE B (continued). On Definition 18, the distances of (*B*1), (*B*2), (*B*3), and (*B*4) from the actual world are 2.57, 2, 0.62, and 0.62, again in full accord with what is intuitively expected.

EXAMPLE C (continued). The distances of (*C*1), (*C*2), (*C*5), (*C*6), (*C*3), (*C*7), (*C*4) from the truth work out, respectively, at 6, 3, 2.73, 1.88, 1.5, 1, and 0.5, thus vindicating all the intuitive judgments of verisimilitude we made concerning those theories in section 1.

In particular, Popper's desideratum that (*C*7), for instance, should have more verisimilitude than (*C*5), is accommodated.

It is somewhat odd that Popper's paradigm example of unequal verisimilitude should feature statements of the type (*C*5) and (*C*7), rather than statements of the form, say, (*C*2) and (*C*4), whose unequal verisimilitude is intuitively much more conspicuous. Popper admits that examples of the latter kind carry considerable intuitive appeal. He writes: "It is now 9.45' seems nearer to the truth than 'It is now 9.40 p.m.' if in fact it is 9.48 p.m. when the remark is uttered'. But he sets this example aside as merely apparent. 'In this form', he explains, 'the intuitive impression is mistaken: the two statements are incompatible and therefore non-comparable (unless we introduce a ***measure*** like *ct*).'[1]

---

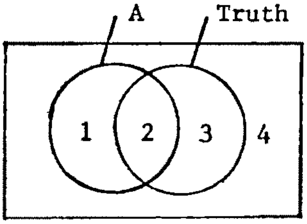
* *Added in proof:* The article has now appeared in *Synthese*, 30 (1975), pp. 159–191.

It is rather difficult to understand why Popper should take the view that incompatible statements cannot be compared as to their closeness to the truth. All the more so because in other places he makes pronouncements which imply the exact opposite of this view. In Popper [1972] he insists in one place[2] that 'Newton's theory contradicts both Kepler's and Galileo's theories . . . and similarly Einstein's theory contradicts Newton's'; and in another place[3] he writes that 'relativity theory is—or so we conjecture—a better approximation to the truth than is Newton's theory, just as the latter is a better approximation to the truth than is Kepler's theory'. If Einstein's theory can be closer to the truth than Newton's despite contradicting it, there seems no reason why 'It is 9.45' should not be closer to the truth than 'It is 9.40' despite contradicting it, and why our 'intuitive impression' on that score should be deemed mistaken.[4]

5 A LOOK BACK ON POPPER'S APPROACH

Popper undeniably deserves credit for highlighting verisimilitude as a concept of prime importance. He does not seem to have appreciated, however, the vast difference in nature which sets this concept apart from the traditional tools of semantic appraisal, especially from the concepts of (absolute) truth and falsehood. Popper's failure to explicate verisimilitude can in fact be traced back to the presumption that the concept can be reduced to, or defined in terms of, truth and falsehood.

Popper seems to approach verisimilitude with the following figure at the back of his mind:



UNIVERSITY

---

[1] Popper [1972], pp. 55–6. [2] P. 16. [3] P. 59,

[4] In a recent letter to the author Professor Popper admitted that he had been wrong in maintaining that incompatible theories were incomparable.

The left-hand circle represents a false theory A and the right-hand circle represents the truth. Region 2, where A overlaps with the truth, pictures what Popper calls A's 'truth content', and region 1 pictures A's 'falsity content'. Now Popper's idea is that A has the greater verisimilitude the greater its truth content and the smaller its falsity content, and that in order to increase the verisimilitude of A one has to add to A's truth content without at the same time adding to its falsity content or to take away from A's falsity content without at the same time taking away from its truth content (or both).

A moment's reflection reveals, however, that this idea won't work if one interprets the circles in the above figure the way Popper does, namely as representing the class of statements following from A and the class of true statements respectively. For suppose we try to improve on A by adding a statement S (from region 3) to its truth content or by removing a statement S' (from region 1) from its falsity content. Now unless the theory thus obtained is true, it will contain a false statement, say F. It is readily seen that by adding S to A's truth content we automatically add $S \,\&\, F$ to its falsity content, and by removing S' from the falsity content we automatically remove $F \supset S'$ from its truth content. Thus every inflation of region 2 brings about a corresponding deflation of region 1, and every deflation of region 1 brings about a corresponding deflation of region 2. On Popper's precept, there is thus no way of increasing the verisimilitude of a false theory short of making it (absolutely) true.

Popper's vision of overlapping circles, however, is not entirely groundless. It can be rescued from the above argument; but the circles must not be thought of as representing classes of statements, but rather classes of *d*-kinds.

Imagine, for simplicity, that A is a consistent *d*-constituent, say C^{ϕ}, and that ϕ^{*} is the *d*-family determined by the actual world. Then the left-hand circle can be taken to represent ϕ and the right-hand circle to represent ϕ^{*}. On this interpretation, Popper's idea can be implemented in a non-vacuous fashion. Taking our cue from Popper, let us speak of the united regions 2 and 4 as the truth-content$_P$ of A and of the united regions 1 and 3 as the falsity-content$_P$ of A. Verisimilitude—call it verisimilitude$_P$—can then be defined, in accordance with Popper's intuitive idea, thus: the verisimilitude$_P$ of A is the greater the greater the truth-content$_P$ of A. (Note that the greater the truth-content$_P$ of A, the smaller, automatically, its falsity content$_P$ and *vice versa.*) Popper's precept concerning increasing the verisimilitude of A becomes non-vacuously true: one can indeed increase the verisimilitude$_P$ of A by adding to its truth-content$_P$ without at the same time adding to its falsity content, or, equivalently, by taking away from its falsity content without simultaneously taking away from its truth content.

EXAMPLE B (continued). Theory (*B*1), we have seen, is equivalent to C^{ϕ_1}. Consider now the theory

(*B*5) $(\exists_1 x)Px \;\&\; (\exists_1 x)Qx \;\&\; (\forall x) \,.\, Px \supset \,\sim Qx.$

It is easily checked that (*B*5) is equivalent to C^{ϕ_5}, where ϕ_5 consists of K_1^3–K_4^3, K_{10}^3, K_{11}^3, K_{13}^3–K_{19}^3. Now all the members of ϕ_5 belong to ϕ^*. Hence the truth-content$_P$ of (*B*5) arises from the truth-content$_P$ of (*B*1) by adding K_{10}^3, K_{11}^3, K_{13}^3–K_{19}^3. Thus (*B*5) has a greater verisimilitude$_P$ than (*B*1). And this is in full accordance with what one intuitively expects. It will be recalled that (*B*1) ascribes to *P* a one-element extension and to *Q* an empty one. (*B*5) ascribes to both *P* and *Q* one-element, mutually disjoint extensions. Thus (*B*5) is patently closer to the truth than (*B*1).

Verisimilitude$_P$ is thus a non-vacuous concept, having intuitively congenial instances. And it seems to me that this concept is as close as one can get to implementing the spirit of Popper's approach.

Yet, as explication of our preanalytic notion of truthlikeness, verisimilitude$_P$ is unacceptable. To see this, consider a simple example.

EXAMPLE B (continued). We have seen that intuitively, (*B*2) is closer to the truth than (*B*1). Yet the verisimilitude$_P$ of (*B*1) is greater than that of (*B*2): the truth-content$_P$ of (*B*1) exceeds that of (*B*2) by $\{K_5^3, K_6^3, K_7^3\}$.

The inadequacy of verisimilitude$_P$ stems from the all-or-nothing attitude it takes to individual *d*-kinds. Given a *d*-kind *K* in ϕ, the only consideration affecting the verisimilitude$_P$ of *A* is whether or not *K* belongs to ϕ^*: if it is in region 2, it contributes to the verisimilitude$_P$ of *A*, and if it is in region 1, it detracts from it. The *nature* of the *d*-kinds in ϕ (*i.e.* the way they distribute truth-values through atomic propositions of *L*) is ignored. What is ignored, more particularly, is how *similar* the members of ϕ are to the members of ϕ^* and *vice versa*. What is missing, for instance, is the insight that if some members of region 1 are very similar to some members of region 3, then it is *almost* as if they *were* in region 3, thus enhancing, rather than detracting from, the verisimilitude of *A*.

EXAMPLE B (continued). The addition of K_5^3, K_6^3, and K_7^3 to ϕ_1 swells the falsity-content$_P$ of the class. Nevertheless, the resulting class ϕ_2 as a whole matches ϕ^* better than ϕ_1 does. To see this, consider, for instance, the member $K^3 = \langle 0,0,1,1,0,0 \rangle$ of ϕ^*. The minimal distance of K^3 from a member of ϕ_1 is 3, while the minimal distance of K^3 from a member of ϕ_2 (*e.g.* from K_6^3) is 2.

To sum up, Popper's idea represented in the above diagram has an element of truth in it. *Other things being equal,* theory A is the closer to the truth the larger its overlap with the truth. This, however, is only part of the story, and by no means the most important part. The main consideration is how closely the members of the left-hand circle can be matched by members of the right-hand circle and *vice versa*. The better the match, the greater the verisimilitude of A.

Assuming that the definition of verisimilitude proposed in section 3 is correct and that verisimilitude is what science aims for, what light is shed on Popper's criteria of comparative excellence and acceptability of theories?

Popper's model of scientific progress is a falsificationist one. According to Popper, a theory A is typically discarded in favour of a competing theory A' when A, but not A' has been observationally or experimentally falsified. 'The theoretician', Popper says, 'will prefer a non-refuted theory to a refuted one provided it explains the successes and failures of the old theory.'[1] In other words, A' is preferable to A if it partakes of A's triumphs and turns some of A's defeats into its own victories.

There is an element of truth in this. Suppose, for simplicity, that both A and A' are d-constituents (rather than disjunctions of such), say C^{ϕ} and $C^{\phi'}$. Imagine that $K_1, \ldots, K_k, L_1, \ldots, L_l$ are all the d-kinds hitherto actually observed and that $K_1, \ldots, K_k$ belong to both ϕ and ϕ' whereas $L_1, \ldots, L_l$ belong to ϕ' only. This is clearly the sort of situation envisaged by Popper: the new theory A' shares the successes of the old theory A and, moreover, succeeds where the old theory fails. Now Popper's advice to discard A in favour of A' can be justified by the following consideration. As all the observed d-kinds belong to ϕ' but only some of them to ϕ, it is rational to conjecture that ϕ' shares more elements with ϕ^* than ϕ does. But we have seen that—other things being equal—the greater the overlap between a d-family and ϕ^*, the greater the verisimilitude of the corresponding d-constituent.

EXAMPLE B (continued). Imagine that (B1) competes with (B3) at a time when the only 3-kinds actually observed are K_1^3, K_2^3, K_{10}^3, and K_{11}^3. We know that (B1) and (B3) are equivalent to C^{ϕ_1} and C^{ϕ_3} respectively. Since ϕ_3 contains all the four known d-kinds whilst ϕ_1 contains only two, it is rational to conjecture that ϕ_3 has more d-kinds in common with ϕ^* than ϕ_1 has. And this in turn provides rational grounds for the conjecture that (B3) has a greater verisimilitude than (B1).

In general, however, the proximity of a d-family ϕ from ϕ^* depends not only on the number of d-kinds that ϕ shares with ϕ^*. It depends on the distance of every member of ϕ from the nearest members of ϕ^* and, conversely, on the distance of every member of ϕ^* from the nearest members of ϕ. If ϕ and ϕ^* overlap, some of those distances are zero; but the remaining, positive, distances also matter. As above, assume that the competing theories A and A' are equivalent to C^{ϕ} and $C^{\phi'}$ and that $K_1, \ldots, K_k$, $L_1, \ldots, L_l$ are all the d-kinds actually encountered. But this time suppose that while $K_1, \ldots, K_k$ belong to both ϕ and ϕ', the remaining d-kinds $L_1, \ldots, L_l$ belong to neither. In this situation A' cannot be said to succeed where A fails, for the adverse observations of $L_1, \ldots, L_l$ falsify A' as well as A. Yet, if the distances between members of ϕ' and members of ϕ^* are on average significantly smaller than those between members of ϕ and members of ϕ^*, A' will be significantly closer to the truth than A. What more, the available evidence may strongly indicate that this is indeed the case: for the average distance of the known members $L_1, \ldots, L_l$ of ϕ^* from members of ϕ' may be significantly shorter than the average distance of $L_1, \ldots, L_l$ form members of ϕ. It seems perfectly rational, in such a situation, to prefer A' to A in spite of the fact that no defeats of A have proved victories for A'. The preference is justified by the fact that some rather disgraceful defeats of A have proved fairly honourable defeats for A'. This, of course, is the logical rationale underlying the common practice of abandoning a theory in favour of a new one whose predictions, although still falling short of correctness, are significantly more accurate than those of the old theory.

---

[1] Popper [1972], p. 14.

EXAMPLE B (continued). Imagine that (B1) competes with (B2) at a time when the only 3-kinds actually found are K_1^3, K_2^3, $K_{20}^3 = \langle 0,0,1,1,0,1 \rangle$, and $K_{21}^3 = \langle 1,0,0,1,0,1 \rangle$. The latter two 3-kinds belong to neither ϕ_1 nor ϕ_2, hence both (B1) and (B2) have been falsified. But note that while the distance of K_{20}^3 and K_{21}^3 from any member of ϕ_1 is equal or greater than 2, their distance from some members of ϕ_2 (*e.g.* K_5^3 and K_7^3 respectively) is 1. It is thus rational to conjecture that the optimal distances between members of ϕ^* and members of ϕ_2 are on average shorter than those between members of ϕ^* and members of ϕ_1, and that (B2) has accordingly more verisimilitude than (B1) has.

Note that ϕ_1 and ϕ_2 share the same 3-kinds with ϕ^*: *viz.* K_1^3–K_4^3. Hence if we appraised the relative merits of (B1) and (B2) in terms solely of (B1)'s falsifications which have proved confirmations of (B2), then, no matter what empirical evidence was available, we would never be led to the conjecture that (B2) is closer to the truth than (B1).

Popper's falsificationist account of scientific change and his notion verisimilitude fail to do justice to the above simple point. Popper's idea to explicate verisimilitude in terms of truth content and falsity content is deficient in failing to discriminate among falsehoods. What is missing is the insight that among the false consequences of false theories, some are less false than others.

University of Otago

REFERENCES

HINTIKKA, J. [1965]: 'Distributive Normal Forms in First-Order Logic', in: J. N. Crossley and M. A. E. Dummet (*eds.*), *Formal Systems and Recursive Functions; Proceedings of the Eighth Logic Colloquium*, Oxford, July 1963, pp. 48–91.

MILLER, D. [1974]: 'Popper's Qualitative Theory of Verisimilitude', *British Journal for the Philosophy of Science*, **25**, pp. 166–77.

POPPER, K. R. [1962]: 'Some Comments on Truth and the Growth of Knowledge', in E. Nagel, P. Suppes, and A. Tarski (*eds.*): *Logic, Methodology and Philosophy of Science*, pp. 285–92.

POPPER, K. R. [1963]: *Conjectures and Refutations.*

POPPER, K. R. [1972]: *Objective Knowledge.*

TICHÝ, P. [1974]: 'On Popper's Definitions of Verisimilitude', *British Journal for the Philosophy of Science*, **25**, pp. 155–60.

Brit. J. Phil. Sci. 27 (1976), 43–50 *Printed in Great Britain*

The Laws behind the Quantum Laws

by ALFRED LANDÉ

Introduction
1 *The Probabilistic Ground Structure.*
2 *Two-way Symmetry and Thermodynamics.*
3 *Relation Between Sets of P's.*
4 *A Geometrical Analogy.*
5 *Correspondence and ψ-Amplitudes.*
6 *Covariance and Quantum Dynamics.*

INTRODUCTION

The mathematician's delight is to see fundamental problems solved by a minimum of logical reasoning (for example, the solution of the problem of whether root two is rational). The scientist's delight is to reduce intricate regularities or apparent irregularities of natural phenomena to simple and general ground principles. This method of 'explaining' odd-looking relations has been gravely neglected in the case of quantum mechanics. When all efforts, during the first decade after Planck's discovery, to reduce the quantum jumps to some gap in classical statistical mechanics proved to be futile, quantum theory became accepted as fundamental, as irreducible to anything more elementary and plausible as a matter of principle. Later in 1926, when a new set of quantal calculation rules was found, hardly any physicist felt the urge to ask for the *Why* of these rules, after their *How* proved so immensely successful.

One of the few exceptions was Albert Einstein. But in his efforts to 'solve the quantum riddle' he was misled by his conviction that 'the Lord does not play dice with the world' for quantum mechanics is an essentially probabilistic structure, incapable in principle of being reduced to deterministic chains (beginning with an undeterministic molecular disorder having taken place five billion years ago?) So the question of 'why are there quantum jumps?' remained unanswered. The reply 'Because there is a Schrödinger wave equation' is quite naive, since this equation contains the various quantum riddles all rolled into one formula.

Received 1 *May* 1975

In the mean time much progress seemed to have been made at least in unifying the various calculation techniques, in particular by von Neumann's *Mathematische Grundlagen der Quantenmechanik*. Unfortunately, this text is so advanced that the library copy, at least at this writer's place, shows heavy use by multitudes of eager readers on the first few pages; then it becomes cleaner and cleaner, to appear as good as new from page 10 on. It therefore remains problematic how much real understanding has been gathered from this learned work which, moreover, deals exclusively with the *How* of the theory, leaving wide open the important question of *why* Nature presents us with such enormous intricacies. As natural scientists and philosophers we rather expect greater and greater *simplicity*, the closer we come to real foundations. Our goal then is to show that the astonishing formalism of quantum mechanics, always accepted as further unexplainable, in fact follows from three simple and plausible postulates, those of *symmetry*, *correspondence*, and *covariance* imposed on the most general probabilistic structure of events. A special restriction for dynamics is, however, that in contrast to classical mechanics, there are no simultaneous states of position and velocity but only separate states of the two—as Zeno maintained 2200 years ago. This *points toward* quantum mechanics but does not contain any hint of quantum mechanics itself with its wavelike connection between position and velocity and the inverse proportionality of their uncertainty regions. Besides, if 'probability' is taken in its objective frequency sense, then there can be only one correct and a handful of almost correct views concerning Measurement, Schrödinger's Cat, and the *EPR*-controversy.

1 THE PROBABILISTIC GROUND STRUCTURE

There may be a variety of observable quantities or 'observables' A and B and C and so forth pertaining to a certain physical object. The observable A may be found in various 'states' $A_1 A_2 A_3 \ldots A_m$. It is left open whether their number or 'multiplicity' m is finite or infinite, and whether the values A_α form a discrete or a continuous set. The same may hold for the observable B with its 'eigenvalues' $B_1 B_2 B_3 \ldots B_n$ and for other observables C and D and so forth.

Suppose an A-measuring instrument or 'A-meter' has ascertained the system in a state A_α. A B-meter test is now applied. The result is unpredictable. But in our probabilistic structure each value B_β has a definite relative frequency or probability to turn up, denoted as $P(A_\alpha \to B_\beta)$ or shortly $P_{\alpha\beta}$. The sum of these probabilities is *unity* of course, that is

$$\Sigma_\beta P_{\alpha\beta} = 1 \text{ for every single } \alpha \qquad (1)$$

There are also the special cases $P(A_\alpha \rightarrow A_\alpha) = 1$ and $P(A_\alpha \rightarrow A_{\alpha'}) = 0$ for $\alpha' \neq \alpha$, abbreviated

$$P_{\alpha\alpha'} = \delta_{\alpha\alpha'}. \tag{1'}$$

2 TWO-WAY SYMMETRY AND THERMODYNAMICS

The general probabilistic structure shall now be restricted by the postulate of a *two-way symmetry* of the P's:

$$P(A_\alpha - B_\beta) = P(B_\beta - A_\alpha) \quad \text{or} \quad P_{\alpha\beta} = P_{\beta\alpha} \tag{2}$$

which allows the initial state to play the same role as the final state in the probability of transition, thus leading from (1) to

$$\Sigma_\beta P_{\beta\alpha} = 1 \quad \text{for every single } \alpha. \tag{3}$$

The symmetry postulate is simple and plausible because it corresponds to the classical reversibility of mechanical processes. Yet its consequence (3) taken as an isolated summation rule would indicate a most surprising fact. It states that a re-arrangement of the P's of (1) into sets all *ending* in one state A_α happen to yield unity, too; and this not only for one final state A_α but for each one of them. The same holds for all other pairs of observables. (This is experimentally confirmed by the intensity distribution in multiple spectral lines).

There is still another argument in favour of (2) besides its correspondence to classical reversibility, namely *thermodynamics* with the Second Law. As shown by von Neumann: When any distribution of like objects over the states A_α is subjected to a B-meter test, the resulting distribution over the new states B_β will in the average have a larger probability and entropy, $S_B \geqslant S_A$, provided that not only the natural rule (1) but also the astonishing summation rule (3) holds. But as remarked before, this rule, and hence the Second Law of thermodynamics, would appear as a coincidence of many coincidences if it did not have one simple and general explanation; the symmetry of the individual probabilities (2).

3 RÉLATION BETWEEN SETS OF P'S

The quantities $P_{\alpha\beta}$ in our scheme must somehow be characterised as *probabilities*, not only by the sum rule (1) for the relation between single P's, but also by the connection between three sets $P_{\alpha\beta}$, $P_{\beta\gamma}$, and $P_{\alpha\gamma}$. This is usually achieved by the addition rule

$$P_{\alpha\gamma} = \Sigma_\beta P_{\alpha\beta} \,.\, P_{\beta\gamma}. \tag{4}$$

The trouble with (4) is that it signifies, on the right hand side, the total probability of arriving from state A_α at state C_γ *via* the set of states B_β. This, however, rarely equals the probability $P_{\alpha\gamma}$ connecting A_α directly with C_γ without intermediate B-states. Thus the relation (4) does not hold when C_γ is identical to $A_{\alpha'}$, for then the left hand side $P_{\alpha\alpha'}$ for $\alpha' \neq \alpha$ is *zero*, whereas the right hand side is a sum of positive products. Nevertheless, since (4) is correct at least in some cases, it can serve as a *model* for an accurate and general theorem connecting the three sets $P_{\alpha\beta}$, $P_{\beta\gamma}$ and $P_{\alpha\gamma}$, assuming that in atomic dimensions *there is a general self-consistent law* connecting the three P-sets in *correspondence* to (4) so as to yield the latter at least in the average.

4 A GEOMETRICAL ANALOGY

Before establishing the desired general theorem of a triangular P-relation between three observables A, B, C, of a system, let us first consider an analogous problem in geometry concerning three points A, B, C in n-dimensional space forming a triangle with sides of lengths L_{AB}, L_{BC} and L_{AC}. The three lengths L do not determine one another, nor are they independent of each other. They rather obey the inequality

$$L_{AB}+L_{BC} \geqslant L_{AC} \geqslant |L_{AB}-L_{BC}| \tag{5}$$

together with $L_{AA} = 0$ and $L_{AB} = L_{BA}$. (5) includes the exceptional case

$$L_{AC} = L_{AB}+L_{BC} \quad \text{(Exception)} \tag{5'}$$

but only when the three points A, B, C are located in this order on a straight line.

We now try to express the inequality (5) by an equation for auxiliary quantities. (Equations always look better than inequalities). But the equation in question must be such as to include also the equation (5′) for the L's though it is valid only in exceptional cases, whereas the desired equation for auxiliary quantities is required to hold always and in general. The way to achieve this condition is to take (5′) as a *model* and introduce quantities, denoted as V, with each V_{AB} connected with the one corresponding L_{AB}, and subject to the equation

$$V_{AC} = V_{AB}+V_{BC}. \tag{6}$$

The V's are known as *vectors*, and the vector equation (6) is indeed made equivalent to (5) when we put

$$L_{AB} = |V_{AB}| \quad \text{and} \quad V_{AB} = -V_{BA}. \tag{6'}$$

Whereas each V_{AB} determines the corresponding L_{AB} uniquely, V_{AB} is determined by L_{AB} only up to its direction. In plane geometry one can express this by the notation $V = L\,e^{i\varphi}$ with a directional angle φ. However, geometry has managed for five thousand years without the use of auxiliary quantities V and complex symbols.

5 CORRESPONDENCE AND ψ-AMPLITUDES

The problem of finding the one and only *general* probability theorem which yields the insufficient ordinary P-law in the average can be solved in an entirely analogous fashion. Take (4) as a *model*, introduce auxiliary quantities $Q_{\alpha\beta}$, *each one* of them somehow connected with the corresponding $P_{\alpha\beta}$ and then let the Q's satisfy the equation

$$Q_{\alpha\gamma} = \Sigma_\beta Q_{\alpha\beta} Q_{\beta\gamma} \quad \text{and} \quad Q_{\alpha\alpha'} = \delta_{\alpha\alpha'} \tag{7}$$

to serve as a general theorem (like the vector law (6), this has a *group quality*, leading from one triple A, B, C to other triples A, B, D; B, C, D; and so forth).

The connection between the auxiliary quantities Q and the probabilities P can now be obtained from the following comparative table:

$$\left|\begin{array}{l|l} P_{\alpha\alpha'} = \delta_{\alpha\alpha'} & \Sigma_\beta P_{\alpha\beta} = 1 = \Sigma_\beta P_{\beta\alpha} \\ Q_{\alpha\alpha'} = \delta_{\alpha\alpha'} & \Sigma_\beta Q_{\alpha\beta} \cdot Q_{\beta\alpha} = 1 = \Sigma_\beta Q_{\beta\alpha} \cdot Q_{\alpha\beta} \end{array}\right|$$

from which follows (we required above that this hold for each $Q_{\alpha\beta}$)

$$P_{\alpha\beta} = Q_{\alpha\beta} Q_{\beta\alpha} = P_{\beta\alpha}. \tag{7'}$$

Hence, if the Q's are supposed to be real, then

$$Q_{\alpha\beta} = Q_{\beta\alpha} \quad \text{and} \quad P_{\alpha\beta} = (Q_{\alpha\beta})^2; \qquad Q_{\alpha\beta} = \sqrt{P_{\alpha\beta}}. \tag{8}$$

In the more general case of the Q's allowed to be complex:

$$Q_{\alpha\beta} = Q^*_{\beta\alpha} \quad \text{and} \quad P_{\alpha\beta} = |Q_{\alpha\beta}|^2; \qquad Q_{\alpha\beta} = \sqrt{P_{\alpha\beta}} \exp(i\varphi_{\alpha\beta}) \tag{8'}$$

with * indicating the complex conjugate.

It is customary to use the letter ψ for Q and to call ψ or Q a 'probability amplitude'. The theorem (7) written for ψ

$$\psi_{\alpha\gamma} = \Sigma_\beta \psi_{\alpha\beta} \psi_{\beta\gamma} \tag{9}$$

is known as the *interference law* of probabilities. It is one of the backbones supporting quantum mechanics, although the quantum h, the action constant connecting dynamical quantities with space and time periods, does not yet occur in it.

For the proof that (7) and (9) together with (8) and (8') lead to (4) *in the average* refer to my [1973].

6 COVARIANCE AND QUANTUM DYNAMICS

The action constant h enters the previous general probability structure through the restriction that dynamics be *covariant*, *i.e.* not dependent on arbitrarily chosen zero points for energy and momentum, nor for time and space coordinates, nor for angular momentum and angular coordinates. The problem is to explain the chief quantum rules which restrict the change (Δ) of energy, of linear momentum p, and of angular momentum

p_φ in terms of periods τ in time, l in space, and φ in angular space, respectively. These rules are:

$$\Delta E = h/\tau, \qquad \Delta p = h/l, \qquad \Delta p_\varphi = h/\varphi. \tag{10}$$

(Examples: An oscillator of frequency $\gamma = 1/\tau$, a crystal with lattice plane distance l, and any body with $\varphi = 2\pi$.) These and other 'quantum rules' of calculation are experimentally confirmed, and form the backbone of atomic dynamics. But they have never been justified on more elementary grounds, hence Einstein's 'quantum riddle'. We contend that they follow from the *postulate of covariance* requiring that physically significant quantities must depend only on differences, E—E', and so forth.

But first we have to introduce a more special feature of the probabilistic mechanics which has not been emphasised enough in previous publications about the subject. This special feature is that, contrary to ordinary mechanics which allows definite (p,q)-combination states, there are only separate states of momentum p and separate states of position q, connected by probability, and the same for energy E and time t. This Zeno-like assumption certainly points toward quantum mechanics; but it is still far away from a quantum mechanics itself, which is characterised by a periodic or *wavelike* probability amplitude connecting p with q and E with t, and by Heisenberg's inverse proportionality of the two uncertainty ranges.

We now can proceed to the promised derivation of the wavelike structure of probabilistic dynamics from the covariance postulate together with the results of the symmetry and correspondence postulates, regarding p-states and q-states as being in a probabilistic relation.

Consider a physical quantity or observable $T(q)$ with q defined with respect to some zero point. We are interested in the average value of $T(q)$ found on the way from an original state p and ending at another state p'. In classical theory it would be defined as

$$\overline{T(q)}_{pp'} = \int P_{pq}\, T(q)\, P_{qp'}\, dq \tag{10}$$

where P_{pq} denotes the transition probability from state p to state q. Within the realm of the probability interference formalism we have often to do with analogous expressions

$$T_{pp'} = \int \psi_{pq}\, T(q) \psi_{qp'}\, dq \tag{12}$$

denoted as a ***matrix element***. If this quantity has any ***physical significance***, no matter what its special meaning is, it should depend only on the difference p-p' for any function $T(q)$. This covariance postulate should apply also to a delta function, $D(q)$ with maximum chosen at any place q. Here (12) reduces to the simple product

$$D_{pp'} = \psi_{pq}\psi_{qp'} = \psi_{pq}\psi^*_{p'q} \tag{13}$$

required to depend on $p-p'$ only. There is only one way to satisfy this requirement, namely by taking for ψ the exponential function

$$\psi_{pq} = U(q) \exp[ip \,.\, \alpha(q)] \tag{14}$$

with the real function $\alpha(q)$ in the exponent. Carrying out a similar consideration with p and q interchanged one arrives at

$$\psi_{qp} = V(p) \exp[iq \,.\, \beta(p)] = \psi^*_{pq} \tag{14'}$$

with the real factor $\beta(p)$. Comparison of (14) with (14′) yields

$$\begin{gathered} U(q) = V^*(p) = \text{constant } C \\ \alpha(q) = qc \quad \text{and} \quad \beta(p) = -pc; \end{gathered} \tag{15}$$

so that we finally arrive at

$$\psi_{pq} = C \exp(ipqc),$$

or when writing the constant c in the form $2\pi/h$

$$\psi_{pq} = C \exp(2i\pi pq/h), \tag{16}$$

containing an action constant h of the dimension of the product pq. Its magnitude in terms of a conventional system of measures can of course not be predicted on general grounds. But the considerations above yield, by way of the three postulates and Zeno's corollary an explanation of the magic function (16) which is the basis of all quantum dynamics as a wave function with wave length $\lambda = h/p$. Corresponding considerations for E and t lead to the wave function

$$\psi_{Et} = C \exp(2i\pi \,.\, Et/h) \tag{16'}$$

and for p_φ and φ to

$$\psi(p_\varphi, \varphi) = C \exp(2i\pi p_\varphi \varphi/h). \tag{16''}$$

From (16) one can derive the other rules of quantum mechanics, the Schrödinger operator rule and his wave equation, the Born pq-commutation rule, Heisenberg's uncertainty rule, and all the rest—as I show in my [1973]. *The quantum riddle is solved.*

Historically: After the original disillusion of not finding elementary reasons for the amazing quantum rules, a long period of resignation followed when physicists were satisfied with having the correct calculation technique, and regarded it as irreducible to any more simple and general principles. Asking for an 'explanation' was considered as senseless, in contrast to Einstein's admonition:

> The supreme task of the theoretical physicist is the search for those general and elementary laws from which one can derive the world picture by pure reason, although the elementary laws themselves cannot be obtained logically but only on a broad view of experience.

In the present instance we hope to have shown, to use the language of Leibniz, that the quantum world is not only 'the best of all possible worlds', but the only possible world under the three general postulates of symmetry, correspondence, and covariance imposed upon a probabilistic relation between 'states' of 'observables' with extra p- and q-states. Therefore, if someone wishes to modify quantum mechanics, he will have to tell us which of the three postulates he intends to change, and how. Questions about the *substance* to which the general theory is to be applied (*e.g.* to subnuclear particles and their fields) offer problems which are on a level far above that of quantum mechanics proper (See Jammer [1974]). This holds already for the Pauli exclusion principle and the spin of a special object, the electron, together with the theory of relativity.

REFERENCES

JAMMER, M. [1974]: *The Philosophy of Quantum Mechanics.* Wiley Interscience Books.

LANDÉ, A. [1955]: *Foundations of Quantum Theory, a Study in Continuity and Symmetry.* Yale University Press.

LANDÉ, A. [1960]: *From Dualism to Unity in Quantum Physics.* Cambridge University Press.

LANDÉ, A. [1965]: *New Foundations of Quantum Mechanics.* Cambridge University Press. Translated into Spanish as *Nuevos Fundamentos de la Mecanica Quantica,* Madrid: Editorial Tecnos, 1968.

LANDÉ, A. [1973]: *Quantum Mechanics in a New Key,* Exposition Press, Jericho, N.Y.

Brit. J. Phil. Sci. 27 (1976), 51–78 *Printed in Great Britain*

Discussions

THE LIMITATIONS OF SINGLE-EXIT MODELS: REPLY TO MACHLUP

1 *Introduction.*
2 *The Compromise Regarding Oligopoly.*
3 *Explanation of Complex Patterns and the Theory of Sellers' Behaviour.*
4 *What does the Firm 'Know'?*
5 *How Rational is the Firm?*
6 *Personal Types and Course-of-Action Types.*
7 *Restatement and Extension.*

1 INTRODUCTION

In the September 1974 issue of this *Journal*, Professor Fritz Machlup discussed some ideas put forward in my 'Situational Determinism in Economics' (Machlup [1974] and Latsis [1972]) where I attempted to apply the methodology of scientific research programmes to the neoclassical theory of sellers' behaviour. My paper contained two main theses: a *descriptive* thesis and an *evaluative* one.

The descriptive thesis may be briefly restated as follows: Most theories of sellers' behaviour put forward from 1838 to the mid-nineteen fifties may be fruitfully viewed as various models of the same dominant research programme. This research programme's 'hard core' was a particular version of what is known in the literature as the assumption of rationality, while the main regulative idea guiding its 'positive heuristic' was that economic behaviour depends on and should be explained in terms of the structure of the economic environment and *not* the decision makers' internal characteristics. More specifically, in explaining economic behaviour, adherents to this programme tend to reduce economic decision situations to 'single-exit' situations, *i.e.*, situations where the uniquely best course of action is narrowly delimited from a consideration of the objective situational constraints.

In the *evaluative* section of my [1972] I argued that a closer look at the relevant area of the history of economic theory reveals that the neoclassical research programme of sellers' behaviour (which I called 'situational determinism'), has to be judged as degenerating, though admittedly there was no rival research programme. It was shown that the various ramifications and elaborations of the neoclassical theory of sellers' behaviour did not lead to empirically testable (let alone corroborated) results. My arguments were based on an appraisal of the logical and empirical adequacy of the theories of monopolistic competition, oligopoly, duopoly and bilateral monopoly.

Machlup's criticism of my [1972] has been helpful to me in various ways. For instance, I have been stimulated into recasting my ideas in different terms which perhaps avoid obscurities and misunderstandings. I have also been led to rethink theses which I had perhaps carried too far as well as theses which I had not carried far enough.

In the present paper I shall proceed as follows: In sections 2 to 6 I shall appraise what I consider to be Machlup's most important criticisms. In section 7 I shall restate and extend the main thesis put forward in my [1972]. This last section will be only an indirect answer to Machlup, but I feel it may help to clear up misunderstandings which obscurities in my [1972] presentation may have caused.

I start by taking up Machlup's most interesting criticisms. I will use (in all but one case), the same headings as he used and will deal with the relevant sections in the same order. Where I am unsure of the location of our disagreement I will do my best to pinpoint it. Where I cannot discern substantial differences I shall endeavour to clarify my own position on the issues in question.

2 THE COMPROMISE REGARDING OLIGOPOLY

Machlup criticised me for affirming 'that I see no prospect for a compromise solution whereby situational determinism retains most of the theory of the firm but concedes oligopoly to behaviouralism'. (Latsis [1972], p. 233). (Machlup seemed to have indicated such a compromise in his [1967]).

Machlup complains that I seemed to be saying that even in cases where the comparatively simple techniques of situational determinism are adequate we should ignore them in favour of the more cumbersome behavioural approach. Would it not be more sensible, he suggests, to use situational determinism where it is applicable and only attempt alternative approaches elsewhere? Quite so. The question is *how* applicable is situational determinism to sellers' behaviour in the variety of market situations?

I objected to Machlup's compromise proposal on the grounds that it could serve to conceal a fact drawn from the history of economic theory, namely, that situational determinism fails to explain and predict sellers' behaviour, except in a *very narrow* range of circumstances. One of the central theses in my [1972] was precisely to argue—*via* a rational reconstruction of the relevant area of the history of economic theory—that the only exemplification of situational determinism that fulfils minimal logical and empirical adequacy criteria is the perfectly competitive model.

It is part of the folklore of neoclassical microeconomics, however, that the theory of sellers' behaviour deals adequately with behaviour in *all* market situations with just one exception: namely, the case where sellers are few, otherwise known as oligopoly.

Another line of defence of the neoclassical theory of competition is that it is not concerned with the explanation of the behaviour of single firms but rather that it is designed to account for the *effects* of behaviour in the aggregate. It is implicit in this defence of neoclassical microeconomics that a model that may give false predictions about the behaviour of individual sellers nevertheless gives correct predictions of the aggregate effects of their behaviour. But if the predictions of the effects of typical individual behaviour are persistently falsified I see no reason for supposing that when the effects are taken in the aggregate the respective predictions will be confirmed by some happy pre-established harmony. If the aggregate consists of five firms and we cannot correctly predict their behaviour (or its effects), individually, how can we expect to do so in aggregate? And the same argument holds if we consider imperfectly competitive industries with a larger number of firms constituted of sub-groups with a relatively small number of firms belonging to each sub-group (for instance, the cosmetics industry).

Another interpretation of this defensive claim might be that the predictions of the aggregate effects of behaviour are of a statistical character. It follows from this that the individualistic explanation schemas which underpin aggregate predictions would either be of the statistical/probabilistic variety or could be dispensed with. But this interpretation is not supported in the writings of neoclassical economists where the explanation schemas are consistently of the deterministic variety. A further and more general interpretation will be discussed in the next section.

In short, I claim that the neoclassical theory of sellers' behaviour explains the allocation of resources and gives answers to a few qualitative questions *only for perfectly competitive market situations.* Once we move away from perfect competition to market situations lying anywhere within the spectrum from imperfect competition to pure monopoly we find that the same questions are either not dealt with or else are tackled in a palpably unsatisfactory way. I tried to show this in my [1972] (pp. 213–22) and my [1974] (pp. 151–259). As these models do not yield coherent price-output solutions it is false to assert that the neoclassical theory of the firm explains allocation when it leaves unexplained price-output determination in all except perfectly competitive markets. That is why I rejected Machlup's compromise.

3 EXPLANATION OF COMPLEX PATTERNS AND THE THEORY OF SELLERS' BEHAVIOUR[1]

In this section I shall attempt to bring forward holistic arguments which lend plausibility to Machlup's conventionalist position towards sellers' behaviour. I doubt however whether either Machlup or Hayek—from whose writings I have drawn heavily in this section—would wish to adopt a holistic approach to economic theory.

In his 'Theory of Complex Phenomena' ([1967]) F. A. Hayek offers a criterion of 'pattern complexity'. According to the proposed criterion a pattern is more complex the greater the minimum number of elements required for it to possess the characteristic attributes of a certain class of patterns. (*Cf.* Hayek [1967], pp. 25–6). If we find it useful to talk about complex patterns—or indeed to solve problems about them—by considering phenomena as organised collectivities (or assemblies) of interrelated things, then we should not be surprised if for certain purposes the properties or values of the elements constituting the pattern are irrelevant in the explanation of its structure and properties. We might for instance, wish to claim that hierarchical structures have, in virtue of the organisation of their constituent elements, distinctive survival characteristics. (*Cf.* Simon [1969], pp. 87–90). The arguments on which such claims are based make no use of the properties or characteristics of the elements as such, but rather invoke the architecture of the relations or linkages *between* the elements of the hierarchical structure (*Cf.* Simon [1969], ch. 4). It is, therefore, conceivable—and often convenient—to explain the properties of a complex whole in a way which is holistic in the above sense.

---

[1] I am grateful to Fritz Machlup for drawing my attention to the more recent writings of F. A. Hayek which I found particularly helpful in redrafting Sections 3 and 7.

According to Machlup, the theory of sellers' behaviour does not (and need not, for the purposes for which it has been designed) seek to explain or predict the behaviour of individual sellers. One interpretation of this view (*Cf. above* section 2), is that the aggregate effects of sellers' behaviour are unaffected by the existence of a few deviant sellers (that is, sellers who do not behave in accordance with the theory of sellers' behaviour we happen to employ). Clearly this depends on the number of deviants and the extent of their deviance. Surely such a theory ought to explain tolerably well the behaviour of most sellers individually as well as in aggregate. This is why in my [1972] paper I proceeded on the assumption that a satisfactory theory of sellers' behaviour must involve at least an approximately correct explanation of how most sellers actually behave. Machlup seems to deny this.

But there is a further possibility which Machlup might wish to endorse: namely, that for certain purposes the aggregate effects of behaviour are independent of individual behaviour. This apparently implausible view may become acceptable on the basis of the holistic view mentioned above. For we may interpret Machlup's position as an application to economics of the observation that under certain conditions the explanation and prediction of the states of complex patterns or changes thereof *may* proceed independently of the properties and values assumed by the multitude of individual elements whose interrelations compose these patterns. It is, therefore, conceivable that certain market phenomena could be independent of the behavioural properties or indeed the behaviour of the sellers in these markets. That is, certain patterns of economic life may repeat themselves irrespective of the regularity of behaviour of their constituent elements. It would be a welcome development if we could successfully predict market phenomena without a theory of sellers' behaviour.

Let me recapitulate: Depending on the questions we wish to answer, complex structures (including economic and social organisations) may be viewed in various ways. One way of viewing them is as organised collectivities of anonymous elements where the properties and values of the elements are essentially irrelevant to the properties of the collectivity. Another way of viewing complex structures is as collectivities of independent elements where the properties of the collectivity are the direct outcome of the constituent elements. A still further way—and perhaps a more sensible one—is to study collectivities both from the point of view of the properties possessed by the architecture of the linkages among the elements and from the point of view of the general properties of the constituent elements. There is no single, correct method of tackling complex structures.

4 WHAT DOES THE FIRM 'KNOW'?

Machlup criticises my claim ([1972]) that the neoclassical theory of sellers' behaviour requires that the hypothetical seller be endowed with 'perfect knowledge' of the relevant situational (especially economic) conditions. This criticism of Machlup's is most interesting in that it touches on an important problem, namely, that of the role of the *degree of information* possessed by a behaving organism—a central problem in the explanation of human behaviour in general and sellers' behaviour in particular.

Machlup asserts that the relevant knowledge is sometimes 'minimal'. He stresses that the only knowledge the imaginary firm needs to be endowed with is the knowledge of 'changes in conditions' ([1974], p. 278), that is, 'the direction in which some pertinent magnitude has changed' ([1974], p. 279). This is surely wrong. Even in simple cases the theory requires the seller to have adequate information about the *shapes* of cost and demand curves. In the neoclassical theories of imperfect competition, bilateral monopoly and limit price theory a considerable amount of information is necessary, sometimes including information concerning the competitors' expected output or the elasticities of cost and demand curves.

When Machlup extols the virtues of the neoclassical theory of sellers' behaviour he is really talking about Marshallian demand and supply and in particular about long-run price adjustments. I agree that for the prediction of the effect of a change in economic conditions on price in the long run (when all adjustments initiated by the postulated change have taken place), we need attribute little precise knowledge to the typical 'seller' and in this sense we can simplify the model drastically. But even this 'minimal' amount of knowledge would exceed information concerning 'changes in conditions', to which Machlup wishes to restrict the seller's knowledge endowment. Otherwise, we could just as well deal with aggregates and forget the individualistic reduction which Machlup seems to favour. However, if we wish to predict the effect on price in the short period or even in the intermediate period, it would be necessary in the neoclassical analysis to attribute to the seller a good deal of knowledge; more than he is likely to have in fact.

Besides criticising Machlup's answer to the question 'What does the firm know'? I would also like to suggest that the question is misguided. I believe that questions referring to the amount of knowledge to be attributed to the typical seller or firm are bound to be misleading because the degree of information a typical agent must be endowed with (in order to behave adequately), depends on his goals and the type of situation in which he finds himself. Under perfectly competitive conditions, in equilibrium, an optimising seller need only know the going price and his marginal cost curve. Under monopolistically competitive conditions the seller must be endowed with more detailed knowledge relating to the demand and cost conditions he faces.

We may conclude by saying that the knowledge endowment of the 'typical seller' in neoclassical theory is a relative matter—relative to the motivation he is endowed with and to the characteristics of the market situation under scrutiny. But the knowledge required by the neoclassical seller must be correct and adequate to the situation. Otherwise the models become indeterminate.

5 HOW RATIONAL IS THE FIRM?

Machlup moves from the question 'how much knowledge should the hypothetical actor be equipped with?' to the question 'how "rational" is the firm?' Both questions are misconceived for very similar reasons. How much knowledge an actor needs will depend, as we have seen, on his aims and on the type of situation he finds himself in. And likewise, how 'rational' a firm needs to be depends on the type of environment in which it operates. Again it is a relative matter and not an absolute one as Machlup seems to think.

I prefer to view rationality as a property attributed to the relation—or interaction—of an organism with its environment. (*Cf.* Simon [1969], pp. 6–9).

Viewed as such, rationality is relative to the aims of the behaving organism, to its environment and to the information the organism possesses about its environment. Rationality is, therefore, a 'plastic' concept—at some times having a richer, at other times a poorer content. For instance, under perfectly competitive conditions, it is sufficient that the seller (to be 'rational'), behaves so as to avoid bankruptcy. Here satisficing rationality is sufficient to yield determinate results. Under monopolistic conditions however, nothing short of optimising rationality will yield determinate results. In intermediate situations the attribution of rationality calls for varying degrees of richness.

To conclude: the question Machlup asks is misguided because of the plasticity of the concept of rationality. My [1972] breakdown of the compressed notion of neoclassical rationality into the various elements of the decision process was meant precisely to illustrate the inherent richness and plasticity of the concept which the methods of situational determinism have served to conceal.

6 PERSONAL TYPES AND COURSE-OF-ACTION TYPES

Machlup takes over the Schutzian distinction between 'personal types' and 'course-of-action types'. The theoretical 'consumer' or 'seller' of neoclassical economic theory is presumably a 'course-of-action type'. Fred, the butcher down the road, is a 'personal type'. Machlup implies that if I had been aware of this distinction I would have made fewer errors.

I am not as appreciative as Machlup apparently is, of Schutz's laborious distinction between various 'types' distinguished by degrees of anonymity and intimacy.

As I see it, the situation is essentially simple: any attempt at theorising, whether in the natural or the social sciences, involves some problem situation and the selection and generalisation of relevent features from the complex variety of phenomena and processes to be found in the subject matter under consideration. Both the theories we employ for explanation, and the things we wish to explain and predict, are expressed in propositions—artifacts of our own choice. It is a trivial consequence of the above that features are left out which could have been incorporated. It is a further trivial consequence that theoretical constructs do not fully represent real actions or real actors, *i.e.* do not describe actual everyday behaviour in its full specificity.

What about varying degrees of 'intimacy' and 'anonymity' which Machlup makes so much of? In my view they simply reflect the degree of selectivity and simplification we have imposed on a theoretical construct standing in for human actors in a social setting. The more selective, the more general and the simpler our model the better, *provided* it fulfils the purpose for which it is designed. The point is what degree of selectivity and simplification can we get away with? I believe that in micro-economics we cannot get away with much except in certain restrictive situations. I offer lengthy arguments to this effect in my [1972] and in section 7 *below*. In his [1974] discussion, Machlup offers no counter arguments drawn from the history of economic theory. And the same is true of those of his other works with which I am familiar. (*Cf.* Machlup [1952] and [1967]). Machlup ends this section with this direct criticism: 'It is a mistake to think that models containing types of greater anonymity yield fully determinate results while models with less anonymous types yield less determinate results.

Any well constructed model furnishes precisly the results the constructor wants it to furnish' ([1974], p. 283). This criticism is based on a serious misunderstanding which I hope to clarify at greater length in section 7. Briefly, it is this. Machlup persists in believing that what separates our views is that he favours 'anonymous theoretical sellers' while I favour 'intimate' models of sellers, applicable from case to case. This misunderstanding was possibly occasioned by the tendency of some behaviouralists to work from case studies without any rigorous theoretical framework. But it is certainly not a position I adopted in my [1972]. On the contrary, I insisted on explanations having as full a generality as is permitted by the type of phenomena we are dealing with.

Would Machlup claim that a 'maximising' model is more anonymous than a behaviouralist 'satisficing' one? Surely not! Both models are equally anonymous though different in their effects if incorporated in models of sellers' behaviour. Consider for instance the theory of monopoly. If we substitute the 'satisficing' assumption for the 'maximising' one then the equilibrial price-output combination is no longer determined. This is the sense in which a different general assumption may render a determinate model indeterminate. Of course, such a 'satisficing' model *could* be made determinate by filling in the appropriate logical gaps. The important question is *not* whether in an explanation the conclusion follows validly from the premises, but rather what sort of propositions the premises consist of, and whether or not these propositions fulfil certain adequacy requirements. It is often very difficult indeed to render a model *situationally* determinate in a satisfactory way. Moreover, an 'indeterminate' conclusion (for instance, the 'contract curve') may follow from premises which fulfil certain adequacy requirements. The traumatic experiences of neoclassical economists with imperfect competition, oligopoly and bilateral monopoly show this vividly, as is explained in my [1972] and [1974].

7 RESTATEMENT AND EXTENSION

The central idea that has regulated the neoclassical programme for sellers' behaviour is as simple as it is elusive. It is based on an Aristotelian approach to the problem of the explanation of behaviour which I shall briefly explain below.

I start by considering the content and form of the explanans. In general, we would expect explanations of behaviour to consist of propositions relating to the behaving organism and of propositions concerning the environment in which the organism finds itself. Behaviour unrelated to any environment is difficult to conceive, and from premises describing states and processes in the outer environment we cannot directly deduce conclusions about human behaviour. It should be noticed that in setting up the problem situation we have already accepted the separation of an organism's 'inner environment' from the 'outer environment' in which it lives and operates. (*Cf.* Simon [1969], pp. 1–22). To be sure, these are very rough categories. We may be talking of the physical, psychological or social environment. Similarly, we may be talking of the physical, psychological, social psychological or social characteristics of the behaving organism's inner environment.

So, in explaining the behaviour, whether of an amoeba, an animal or a human, the explanans should make some reference to the *internal structure and organisation* of the behaving organism, and should contain some account of the structure and organisation of the *environment* in which this organism finds itself. And if our concern is with an Aristotelian interpretation of human behaviour we might add *aims* and *objectives* not apparent from the study of the internal organisation of the organism: we might consider attributing to the organism aims, preferences or desires which are distinct from its physiological needs or psychological urges and are not traceable back to them.

Some behaviour may be said to be predominantly artificial or conventional in the following sense: it can be understood in terms of an organism seeking a conventional goal and adapting to a largely 'man-made' environment. The burden of explanation is in such cases born by the description of the artificial environment and there is almost no reference to the structure and properties of the behaving organism. (*Cf* Simon [1969], ch. 1). The single-exit approach to explanation and prediction in neoclassical microeconomics is one manifestation of the recognition that in certain areas such artificial explanations are both possible and adequate. The point about single-exit situations is that we need to know little if anything about the behaving organism's inner structure and organisation. It is sufficient to know its aims and the organisation of its relevant environment to understand and to explain its behaviour. Consider the behaviour of a motorist braking in order to avoid collision with the car in front of him. Provided the situation is so structured that no other avoiding action can be taken, the subsequent behaviour is 'situationally determined'. Or, to put it another way, the behaviour is *predominantly artificial* in the sense that it reflects the structure and properties of the environment rather than the properties of the behaving organism. We need to know very little about the internal physical and psychological structure of the driver to predict that he will brake sharply. All we need is some knowledge of his immediate outer environment and the assumption that he wants to avoid collisions. Drivers with very different physical ability, with different characters and attitudes towards risks would take very similar action when faced with the same circumstances. Under certain circumstances a parachutist's legs and the suspension of a motor-car behave in the same way. The reasons for this similarity in behaviour are not to be found in the details of the structure of each of the behaving systems but rather in the *goal* and in the *relevant environment*. Needless to say, the interpretation of behaviour as 'artificial' simplifies the explanatory task when we are dealing with organisms whose internal complexity is considerable.

The possibility of providing simple and effective explanations by viewing behaviour as artificial does not of course imply that reduction to deeper explanations should not be attempted. Nor does it imply that we have unearthed a major difference between the subject matters of the natural and the social sciences. For instance, a thermostat's behaviour may be understood in terms of a 'purely artificial' schema in the same way that a seller's behaviour can, within a certain range of conditions. If the conditions change beyond the appropriate range then the attribution of artificiality may not be as helpful. In the case of a thermostat it is *not* difficult to envisage such conditions. If the thermostat is subjected to variations in temperature beyond its range of tolerance, then its behaviour (as well as the effects of its behaviour) will not be explicable merely on the basis of a goal and of the relevant environment.

In the research programme critically examined in my [1972] the inner environment of the decision maker and the organisation of his decision-making have little or no role in determining economic behaviour. Rather, it is the structure and organisation of the *economic obstacles* (costs, technology, demand) in his outer environment that, given the decision maker's goals, determine his behavioural path. In economics we are rarely if even concerned about the limitations of the economic agents. We assume that the behaving organism has an unlimited capability for optimal adaptation.

And this approach extends beyond economics to the explanation of problem-solving behaviour. Simon writes, in connection with problem solving paths: '. . . it is important to notice that, once the strategy was selected, the course of the search (for a solution) depended on the structure of the problem, not on any characteristics of the problem solver'. To see this, consider a rather simple and straightforward proof in geometry. Suppose we select as our strategy 'analysis-synthesis'. We assume that we have proved the theorem and work backwards to something given in the construction of the problem. We then reverse the process, thereby obtaining the proof. We would be surprised if the search for the proof of an easy problem using 'analysis-synthesis' varied significantly between students of varying behavioural properties. We would expect on the contrary that the paths leading to the proof would be determined by the structure of the problem. But consider now the attempt to solve a very difficult (or even a moderately difficult) geometrical problem. Here we would expect to encounter a considerable degree of failure manifested by unsuccessful search paths. Another thing we would be likely to find is a *variety* of unsuccessful search paths. This variety does not mean that there are no regularities in unsuccessful search paths. Merely, that the differences among the various unsuccessful search paths *cannot* be attributed to the structure of the problem situation. The obvious place to turn to for explanatory causes are the inner characteristics of the problem-solver.

These observations allow us tentatively to formulate a useful and common-sensical criterion for applicability of single-exit models for the explanation of behaviour: *The applicability of single-exit methods of analysis to the explanation of behaviour varies* inversely *with the complexity of the behaving organism's goals and relevant environment in which it operates and* directly *with the amount of information the organism possesses about this envirnoment.*

It follows from this criterion of applicability that whenever the decision maker's (economic or other) cognitive limits are stretched, single-exit methods of analysis are unlikely to be fruitful. That is to say, if an environment stretches an organism's adaptive capacity to its limits—be they physiological, psychological or cognitive—then subsequent behaviour cannot depend on the environment alone. The 'artificial' or externalist explanation of behaviour must give way to a more 'natural' or internalist one. Emphasis here is of primary importance. I am not suggesting that we should adopt purely psychologistic explanations, let alone purely physicalist ones. But I fear that Machlup is suggesting that we should interpret sellers' behaviour in a wide range of market situations as behaviour which is highly 'artificial' in the above sense. The main thesis of my [1972] is that, following an appraisal of its theoretical development it transpires that this explanatory effort has failed. In the final section I have tried to indicate some reasons (however tentative) why we should have expected it to fail. It fails because, for explanatory purposes, it gets the balance wrong between 'artificiality' and 'naturalness'.

SPIRO J. LATSIS
Harvard University

REFERENCES

FELLNER, W. [1949]: *Competition Among the Few.*
HAYEK, F. A. [1967]: *Studies in Philosophy, Politics and Economics.*
LATSIS, S. J. [1972]: 'Situational Determinism in Economics', *The British Journal for the Philosophy of Science,* **23**, pp. 207–45.
LATSIS, S. J. [1974]: *Situational Determinism in Economics.* Unpublished Ph.D. Thesis. University of London.
MACHLUP, F. [1952]: *The Economics of Sellers' Competition.*
MACHLUP, F. [1967]: 'Theories of the Firm: Marginalist, Behavioural, Managerial', *The American Economic Review,* **57**, pp. 1–33.
MACHLUP, F. [1974]: 'Situational Determinism in Economics', *The British Journal for the Philosophy of Science,* **25**, pp. 271–84.
SIMON, H. A. [1969]: *The Sciences of the Artificial.*

ON URBACH'S ANALYSIS OF THE 'IQ DEBATE'

1 Peter Urbach's recent article 'Progress and Degeneration in the "IQ Debate" '[1] is welcome insofar as it constitutes an attempt to analyse the scientific content of a continuing, and (as he rightly states) acrimonious, controversy in the social sciences. It also advances many penetrating criticisms against the views of individual protagonists in the current debate. Urbach's attempt to provide clearly defined standards for the assessment of rival theories and his efforts to detach the scientific debate from its political and emotional concomitants also make the article a significant and a timely one.

Nevertheless, Urbach's own analysis is open to a number of criticisms. This note aims to deal with two of these in some detail. I believe that other aspects of the article are also invalid or misleading, but the two matters raised here are so vital as to raise grave doubts as to the validity of much of his analysis.

2 In the first place, much of Urbach's discussion and a large section of his reference list is devoted to the debate initiated by the writings of Professor Arthur Jensen.[2] Jensen has claimed in essence that the low IQ scores found among *U.S.* Negroes are genetic rather than environmental in origin. This contention has been denied by various scientists representing different branches of specialisation in the biological and social sciences.

Rather than discuss the 'private, psychological beliefs' of scientists, which Urbach rightly states to 'have no place in an appraisal of their work', he essays to consider 'the heuristic principles which they follow *in practice*'.[3] To this end, he states two 'hard cores':

---

[1] Urbach [1974].

[2] Jensen [1969].

[3] Urbach [1974], p. 109: Urbach's italics.

(*H1*) *'All individuals possess a general mental capacity called "general intelligence" which enters with some (and varying) degree into all the diverse types of cognitive activity.'*

(*H2*) *'Differences between individuals and between groups in "general intelligence" are the results of inherited differences.'*

and

(*E1*) *'All individuals inherit all sorts of different mental capacities but each is identical in each individual; all differences in cognitive abilities are the results of environmental differences.'*[1]

It should first be pointed out that these positions by no means exhaust the available possibilities. Espousal of propositions (*H1*), (*H2*), the 'hard core of the hereditarian programme', necessarily entails rejection of proposition (*E1*), the 'hard core of the environmentalist programme', and *vice versa.* However, it is perfectly possible to describe other 'hard cores' that are incompatible with both of those detailed by Urbach.

In particular, a third possible programme, which Urbach neglects entirely, has a 'hard core' which could be expressed as:

(*A1*) *All individuals inherit all sorts of different mental capacities but these do not differ significantly between human groups; all* intergroup *differences in cognitive abilities are the results of environmental differences.*

(Note that this 'hard core' is quite different from any we could write for the 'interactionist' programmes criticised by Urbach.[2])

The complete neglect of proposition (*A1*) is particularly unfortunate in the context of the present debate. It is implicit in the work of many of Jensen's critics, among them the geneticists Bodmer,[3] Crow[4] and Lewontin.[5] These authors are all at some pains to point out that considerations of *intragroup* heritability do not allow any conclusions to be drawn as to *intergroup* heritability. Jensen's argument depends upon the validity of such an extrapolation, and indeed Jensen himself recognises this,[6] although Urbach, who reports that 'Jensen is clearly impressed by this argument',[7] does not.

Particularly relevant in this connection are the three *gedankenexperimente* of Lewontin's first paper.[8] The entire purpose of these is to render plausible the acceptance of proposition (*A1*) in the face of Jensen's extrapolation from familial data. Urbach's hard core statements, with their neglect of the distinction between intergroup and intragroup differences, allow as confirmatory evidence for the hereditarian view the familial data summarised in his Section 2. His account of the state of current debate is thus highly misleading.

The false dichotomy between (*H1*), (*H2*) on the one hand, and (*E1*) on the other has other unfortunate consequences. Jensen is cast as an hereditarian, so that Urbach sees Crow, who expresses some doubts as to the validity of Jensen's arguments, as an environmentalist. Urbach supports this description by stating that Crow believes 'that the interpretation of the Negro-Indian differences in IQ as genetic in origin is unacceptable since 'some environmental variable associated with being black is not included in the rating" '.[1]

---

[1] *Ibid.*, p. 102: Urbach's italics.

[2] *Ibid.*, pp. 110–11.

[3] Bodmer [1972].

[4] Crow [1969].

[5] Lewontin [1970*a*] and [1970*b*].

[6] Jensen [1972], p. 162, note.

[7] Urbach [1974], p. 243.

[8] Lewontin [1970*a*].

Crow, however, wrote[2]: 'I agree for the most part with Jensen's analysis. Any differences could probably best be described by saying that, in general, I have somewhat less confidence than he in the quantitative validity of his methods—more reservations about the reality of the necessary assumptions. I do not mean by this that I would reach opposite conclusions; I am simply more agnostic', and later[3] 'the argument [on Negro-Indian differences] . . . can and will be dismissed on the . . . grounds [that] some environmental variable associated with being black is not included in the environmental rating'.

It is perfectly reasonable for Urbach to discuss the heuristic actually adopted by any scientist, rather than the position he claims to expouse. However, in *this* area, and indeed in others, he does not really do so. I find no indication whatsoever that Crow's work assumes proposition (*E1*), and (beyond the blatant misrepresentation of the out-of-context quotation) Urbach offers no evidence on the point.

3 Urbach's critique of the environmentalist programme (his Section 3) derives much of its force from its characterisation of that programme as leaning heavily on the use of *ad hoc* hypotheses. I believe that much of this section is naive in that it mistakes the nature of the objections to Jensen's work and wrong in that it fails to recognise quite standard and basic scientific distinctions.

Urbach rightly states[4] that 'the hereditarian programme makes no predictions concerning the relative average intelligence of different racial and social groups'. Jensen, for example, does not predict; he offers the hereditarian explanation *ex post facto*. It is the adequacy of this explanation that is in dispute.

Many of Jensen's critics have been content to contest the validity of his arguments—they have not (*pace* Urbach) advanced rival theories of their own. Thus, Crow (whose case is discussed above) is said to appeal to an *ad hoc* hypothesis, when his concern is to make a methodological point. The same discourtesy is done to Medawar,[5] Deutsch[6] and Rex,[7] among others.

---

[1] Urbach [1974], pp. 241–2.

[2] Crow [1969], p. 302.

[3] *Ibid.*, p. 308.

[4] Urbach [1974], p. 235.

[5] *Ibid.*, p. 242. Medawar's point (Medawar [1974]) is essentially that many non-linear phenomena are misleadingly analysed if one relies, as Jensen does, on the (entirely linear) statistical theory of correlation. In keeping with the style of the 'popular' journal in which he writes, Medawar expresses his objection in popular language. It is silly and misleading to represent his example as an hypothesis or to dismiss it (as Urbach does) as an 'empty verbal quibble'.

[6] Urbach [1974], p. 241. Contrast the methodological (although trenchant) tone of Deutsch's paper (Deutsch [1969]), especially the section entitled 'Extrapolations, Contradictions and Misinterpretations'.

[7] Urbach [1974], p. 253. Rex sins slightly by calling the Negro history of slavery the "*crucial* variable" (Urbach's italics). Had he called this history "an *important and neglected* variable", there would have been no possible quibble. Note in this connection a further Urbach inaccuracy. Rex on p. 170 (not p. 87, as stated in note 4 of Urbach's p. 253) regards it as "fortunate" morally that controlled experiment is impossible. This is quite different from Urbach's contention that he regards this impossibility as " 'fortunate' for the hereditarian programme". *Vide* Rex [1972].

Urbach is rightly concerned[1] to 'detach the debate from the political positions with which they are falsely associated in the public mind'. It is, however, unfortunate that in pursuance of this endeavour, he deems it proper entirely to neglect the political context in which much of the debate has taken place. Jensen, for example, is very much concerned with the direction of public policy in the U.S.A. 'Compensatory education has been tried and it apparently has failed.'[2] The issue is one which can hardly fail to evoke emotional and political response and indeed it has done so.

Many of the authors quoted by Urbach write quite unashamedly within such a context, and their main purpose in writing is to demonstrate weaknesses in Jensen's thesis. The debate has been conducted in many cases in 'popular' journals. Although Jensen himself makes the 'IQ Debate' a major part of his research, many of the others quoted do not, but write rather in response to direct editorial invitation or to what they feel to be the dictates of duty. It is neither accurate nor realistic to represent such writing as the statement of scientific hypothesis, as Urbach does.[3] A similar point emerges from an examination of the style of some of the contributions to the debate. To give a particularly clear example: Cottle[4] presents (in literary rather scientific language) a powerful argument that these studies do violence to the social realities they purport to describe.

Among the more cogent scientific criticisms of Jensen's thesis are Lewontin's *gedankenexperimente*, Medawar's methodological points[5] and the question raised by Crow,[6] Lewontin[7] and particularly Rose,[8] who point out that (particularly in human populations, where 'cultural inheritance' obtains) non-genetic mechanisms may mimic genetic ones. In this last context, Urbach overlooks an important scientific distinction.

When geneticists distinguish heredity from environment, they use the term 'heredity' to refer precisely to those mechanisms of familial similarity which are the subject of Mendel's laws, their later extensions and their subsequent cytological and biochemical elucidations. Crow[9] and Lewontin[10] both point out that the existence of other mechanisms of familial similarity as well as these gives rise to systematic overestimates in the (ineptly named) heritability coefficients. Rose,[11] in particular, has described some of these other mechanisms. Urbach[12] objects to allegedly *ad hoc* aspects of Rose's work, which is unfair as Rose[13] describes experiments which potentially allow us to estimate the 'indeterminate distance' to which Urbach objects. Particularly unjust and incorrect is Urbach's characterisation of Rose[14] as 'toppl[ing] over into the hereditarian camp'.

The charge that environmentalists depend increasingly on *ad hoc* hypotheses thus neglects many important considerations. Beyond these, it may also depend crucially upon Urbach's almost exclusive use of archetypes from physical science.[1] This, however, is a difficult question, and I do not pretend here to embark upon a complete discussion.

---

[1] Urbach [1974], p. 100.

[2] Jensen [1969], p. 2.

[3] See in this connection p. 62, note 5 *above*. Urbach's reference list includes articles in (*inter alia*): *The Atlantic Monthly*, *Saturday Review*, *New Statesman*, *New Society*, *The Times Higher Education Supplement* and *New Humanist*.

[4] Cottle [1969].

[5] *Vide* p. 62 note 5, *above*.

[6] Crow [1969], p. 307.

[7] Lewontin [1970*b*], p. 24.

[8] Rose [1972], pp. 141–3.

[9] Crow [1969], p. 307.

[10] Lewontin [1970*a*] and [1970*b*].

[11] Rose [1972], p. 142.

[12] Urbach [1974], p. 253.

[13] Rose [1972], p. 142.

[14] Urbach [1974], p. 253.

4 Although Urbach wishes[2] to free the scientific component of the debate from its political overtones, this is a matter more easily said than done. Because the debate already has such an aspect, any contribution, even the most academic, will be (perhaps undeservedly) judged in these terms. In this context, and also even in terms of pure methodology, Urbach is remiss in not giving space to a proposed experiment, the proposal of which is the concern of many of those whom he characterises as 'environmentalists'. It is this: Suppose the U.S. Government were to embark upon an unequivocal commitment to Negro advancement on all fronts (*e.g.* housing and employment, as well as education). In these circumstances, one might find after (say) three generations a diminution in the admitted and disturbing difference between black and white IQ's?

Besides being an honourable course of action, this is also a perfectly feasible experiment. Its implementation would go far towards deciding between the various schools of thought in this matter.

MICHAEL A. B. DEAKIN
Monash University and
University of Exeter

REFERENCES

BODMER, W. F. [1972]: 'Race and IQ: The Genetic Background', in K. Richardson, D. Spears and M. Richards (*eds.*): *Race, Culture and Intelligence*, pp. 83–113.

COTTLE, T. J. [1969]: 'The Politics of Pronouncement: Notes on Publishing in the Social Sciences', *Harvard Educational Review*, 39, pp. 301–9.

---

[1] It is unfortunate, if understandable, that Urbach's examples of what a scientific theory should look like are drawn almost exclusively from the physical sciences. It is at least arguable that the non-scalar nature of the variables (alternatively describable as the large number of scalar variables) involved alone constitutes an important difference. IQ, for example, is measured as a scalar, although the factor analysis used to set up the notion of 'general intelligence' assumes it to be a vector quantity. Furthermore, *g*, the general intelligence factor, is itself by regarded Jensen [1972] as a 2-vector in his later writing.

The argument is even more peculiar in respect of environment. This is clearly not a scalar variable, although it is sometimes measured as such. Urbach [1974], p. 241, note 2, mentions a list of twelve 'components' of environment. This is intuitively impressive, but we have no idea that the list is exhaustive, or if its components are independent. The *a priori* likelihood is that correlations exist between the components and that other elements are neglected.

The whole question of measurement in the social and psychological sciences depends, in a way that measurement in the physical sciences does not, on statistical technique—in particular the theory of correlation. This theory *assumes* an underlying linearity, which is intuitively implausible, at least over large ranges of parameter values. The thrust of many applications of 'Catastrophe Theory' to the social sciences has been to demonstrate the severe limitations of the linearity assumption.

There are also other possible grounds on which one might argue an essential difference of methodology. It is much harder in the social sciences to disassociate the 'private, psychological beliefs' of the experimenter from his results than it is in the physical sciences. Moreover, particularly in an emotionally charged context such as the 'IQ Debate', it is difficult not to be aware of the immediate social and political consequences of one's conclusions. These consequences do not disappear simply because one prefers not to notice them.

[2] Urbach [1974], p. 100.

CROW, J. F. [1969]: 'Genetic Theories and Influences: Comments on the Value of Diversity', *Harvard Educational Review*, **39**, pp. 301–9.

DEUTSCH, M. [1969]: 'Happenings on the Way Back to the Forum: Social Science, IQ, and Race Differences Revisited', *Harvard Educational Review*, **39**, pp. 523–54.

JENSEN, A. R. [1969]: 'How much can we boost IQ and scholastic achievement?', *Harvard Educational Review*, **39**, pp. 1–123.

JENSEN, A. R. [1972]: *Genetics and Education.*

LEWONTIN, R. C. [1970*a*]: 'Race and intelligence', *Bulletin of the Atomic Scientists*, **26**, pp. 2–8.

LEWONTIN, R. C. [1970*b*]: 'Further remarks on race and the genetics of intelligence', *Bulletin of the Atomic Scientists*, **39**, pp. 23–5.

MEDAWAR, P. E. [1974]: 'More Unequal than Others', *New Statesman*, 11 January.

REX, J. [1972]: 'Nature versus Nurture: The Significance of the Revived Debate', in K. Richardson, D. Spears and M. Richards (*eds.*): *Race, Culture and Intelligence*, pp. 167–78.

ROSE, S. [1972]: 'Environmental Effects on Brain Behaviour', in K. Richardson, D. Spears and M. Richards (*eds.*): *Race, Culture and Intelligence*, pp. 128–44.

URBACH, P. [1974]: 'Progress and Degeneration in the "IQ debate" ', *The British Journal for the Philosophy of Science*, **25**, pp. 99–135, 235–59.

THE MUTE SELF: A REACTION TO DEWITT'S ALTERNATIVE ACCOUNT OF THE SPLIT-BRAIN DATA

Introduction.
1 *Language in the Minor Hemisphere.*
2 *Mirror Images and Plastic Cut-Outs.*
3 *Behavioural Evidence for Self-Awareness in the Minor Hemisphere.*
4 *The Minor Hemisphere on its Own.*
5 *Mute Person or Conscious Part of a Person's Body?*
6 *The Relative Importance of Persons.*

INTRODUCTION

Dewitt ([1975]) offers us an alternative account to previous interpretations of the split-brain studies, one which he represents as a compromise solution between varying views. In brief, his view is that these studies show the split-brain subject has two conscious minds organised in the disconnected cerebral hemispheres, but that only one of these—the major or speaking hemisphere's mind—is self-aware and thus fully qualified as a *person*. Moreover he agrees with me (Puccetti [1973]) that what is found here reveals the true state of affairs in normal, cerebrally-intact humans. But whereas I held that it shows that there are two persons per human body, Dewitt says, of course, that the studies reveal one person and two minds.

A word of caution is necessary at this point. It is true that Eccles and MacKay have held there is only one person and one mind—though for different reasons, as Dewitt explains. And it is true that Sperry and Gazzaniga are willing to countenance two minds in a single split-brain human. But neither of these gentlemen, so far as I know, accepts the inference from two minds in the bisected brain to two minds in the normal (*cf.* Sperry [1974], p. 15). Only Bogen among the principal investigators of split-brain phenomena does so, and it would be unfair to the others to leave this disagreement unstipulated.

However, since Dewitt has joined Bogen and me in making that inference, I will concentrate on his reasons for holding that the minor hemisphere, though constituting the basis for a conscious, intelligent mind, fails to make it as a person even in the brain-bisected human.

1 LANGUAGE IN THE MINOR HEMISPHERE

Throughout his essay, Dewitt assumes that the minor, non-dominant (for speech and handedness: in about 90 per cent of well-lateralised adults the *right*) cerebral hemisphere is devoid of language abilities, which are according to him concentrated exclusively in the major, dominant (usually the *left*) cerebral hemisphere. I shall overlook here the complication that in adult females, sinistrals of either gender and ambidextrals there is a tendency towards bilateralised speech functions; and will dismiss also anomalies such as the girl with agenesis of the corpus callosum who had full speech representation in both hemispheres (Saul and Sperry [1968]). In biology there are a lot of variations and we have to be content with what is preponderantly the case.

But a little reflection will show that the minor hemisphere cannot be entirely without language, since in the commissurotomy experiments one is able to elicit intelligent responses from it using verbal or written instructions (*cf.* Sperry [1968]). 'Draw the figure you see with your left hand', for example, leads to the left hand drawing a figure projected tachistoscopically in the left visual half field when nothing, or a different figure, is flashed in the right half field: and we *know* the minor (or right) hemisphere did it, given decussation of the visual pathways and motor control plus the circumstances of the testing situation. One could multiply this instance a hundred times from the literature.

So the minor hemisphere, while it does not *speak*,[1] certainly *understands* human language. How much does it understand? The early studies of minor hemispheric language comprehension showed it severely deficient compared to the major hemisphere (*cf.* the excellent review by Nebes [1974]), especially for written language involving verbs and whole sentences, but more recent work indicates that the mute hemisphere is capable of extensive *auditory* language comprehension including complex semantic and syntactic abilities (Zaidel [1973]). The minor hemisphere can manifest an auditory picture vocabulary equivalent to that of a 16-year-old normal (Zaidel [1974]).

Is this language ability sufficient, on Dewitt's criterion, for self-awareness?[2] Can the disconnected mute hemisphere which responds to verbal directions containing the personal pronoun 'you' (as in 'Now I want you to reach behind the screen with your left hand and find the object you were holding out of sight before') *fail* to be self-aware? Oddly, the only evidence Dewitt adduces against such an interpretation comes not from the split-brain studies but from experiments with chimpanzees!

---

[1] Even this needs to be qualified. The minor hemisphere in well-lateralised adults is capable of some automatic speech, *e.g.* giving or writing the individual's name, expletives, singing the words of a familiar song once it gets started by others, *etc.* (*cf.* Smith [1966] and discussion below).

[2] Note that no one has shown *any* language ability is necessary to self-awareness. For all we know the concept of self as a subject of continuing experiences antedates formation of a verbal self-concept in the human child and could, therefore, exist in higher animals that are not language-using animals at all. No doubt we feel more confident in ascribing a self-concept to entities that give verbal evidence of its possession, and no doubt the notion of a nonverbal concept of self is phenomenologically barren for us, since we process our experiences in a verbal conceptual scheme. But the existence of such a concept remains very much an open question.

2 MIRROR IMAGES AND PLASTIC CUT-OUTS

Dewitt draws our attention to a study in which chimpanzees were shown able to respond to mirror images of themselves as representing their own bodies and faces rather than the bodies and faces of other chimps, whereas monkeys were not able to make this distinction (Gallup [1970]). He then links this to another experiment in which it was shown that a female chimpanzee could learn symbolic language functions including the construction of whole sentences nonverbally, using plastic cut-outs as arbitrary symbols for words (Premack [1971]), again something monkeys seem unable to do. According to Dewitt, these facts '... would lend corroborating support to the conjecture that language is essential for self-awareness' (p. 44). He then goes on:

> Now I certainly believe that the minor hemisphere lacks anything like full human self-awareness; I suspect, for example, that it does not agonise over its inevitable death the way its companion hemisphere does. In this sense, then, I am inclined to say that it is less than human. Whatever the case may be, we can never ascertain whether the minor hemisphere is deficient in this manner, owing to its muteness. However I am willing to conjecture that not only does the minor hemisphere fail to achieve full human self-consciousness, but that it even fails to achieve the level of self-awareness exhibited by the higher apes! Since this level of self-awareness is apparently testable through purely nonverbal behavioural evidence, the mute hemisphere may have a fair chance of evincing whatever level of self-consciousness it may have (p. 44).

One hardly knows where to begin here. The two experiments Dewitt cites do not provide corroboration of his conjecture that language is essential to self-awareness, but only that the brain of higher apes, and not that of monkeys, is large and complex enough to do two things: one, learn to recognise one's own body and face in a mirror; and two, learn some symbolic language functions nonverbally. The co-existence of two higher cognitive abilities in one animal certainly does not support the view that one *depends* on the other!

Furthermore, where language-learning is concerned, we have already seen that the minor human hemisphere can and does quite naturally learn to understand a good deal, at least, of spoken language. By contrast the chimpanzee, while showing cognitive abilities preadaptive for language acquisition, does not develop this in natural conditions and does so under human tutelage only with arduous training and limited success (*cf.* the exhaustive review by Passingham and Ettlinger [1974]). Indeed in human beings with *global* aphasia, *i.e.* where all language functions in both hemispheres have been destroyed by cerebrovascular accident, symbolic functions can be re-learned far more quickly and easily than chimpanzees can learn them, *using the same plastic cut-outs* (Glass, Gazzaniga and Premack [1973]).

Dewitt by the remarks quoted above seems to think the minor hemisphere would prove less able to recognise its own face and body than a chimpanzee; indeed he urges construction of an experimental design to test his claim (p. 44).

This should be quite easy: one need only flash a series of photographs tachistoscopically in the left visual half field of a right-handed split-brain patient, asking him to press a button with the left hand when he recognises himself. I have no doubt about the outcome. In fact I would venture that if one made the test difficult enough, *i.e.* by mixing into the series many photographs of faces similar to the subject's, the minor hemisphere would do significantly *better* than the same patient's major hemisphere, which Dewitt believes alone capable of facial and bodily self-recognition. The reason I feel confident about this is that a decided minor hemispheric superiority for facial recognition (though other, unfamiliar faces) over the speaking hemisphere of the same split-brain patients has *already* been demonstrated (Levy, Trevarthen and Sperry [1972]). Such a result would also seem predictable from clinical evidence that failure to recognise faces, including one's own face in a mirror or photograph, most often arises from a lesion in the *right* parietal lobe (Hécaen and Angelergues [1962]).

3 BEHAVIOURAL EVIDENCE FOR SELF-AWARENESS IN THE MINOR HEMISPHERE

Of course it does not follow from the fact that the minor hemisphere has some language comprehension and can recognise its face or body in a mirror that it is self-aware. Dewitt may have asked too little here as evidence of self-consciousness in the mute hemisphere. What we need particularly are demonstrations that in the split-brain subject, at least, the disconnected minor hemisphere is capable of showing awareness of self *as distinct from* the dominant, talkative major hemisphere. I gave some indications of this ability in my [1973], where the literature describes the patient's face frowning and his head nodding negatively at each wrong answer produced by the speaking hemisphere about, for example, what object is being palpated out of sight in the left hand.

Bogen ([1974]) has recently provided an even more striking example. In a similar situation, where the test item in the left hand was in fact a pencil, the left hand after several wrong guesses by the speech hemisphere slowly rotated the test object until its point was against the thumb of that hand and then pressed the thumb down hard. Immediately the speaking hemisphere said 'Pencil!', followed by a smile and affirmative head-nodding. What happened here? Well, two facts are especially relevant. First, the pencil was the only pointed object among the array of fifteen or so familiar test items being used. Both hemispheres could be expected to know this, and the speaking hemisphere had only to run through the list of remembered objects to pick out the pencil. Second, pain is perceived bilaterally, *i.e.* in both hemispheres, even in distal parts of a limb on one side of the body (light touches are also perceived bilaterally, but only in the head and neck region). This the mute hemisphere could have learned, not by studying neuroanatomy, but from the experience of hearing the speech hemisphere complain of injuries to the left hand or foot, and also by feeling itself painful stimuli afflicting the right side of the same body.

But then how could the minor hemisphere *not* be aware of itself as distinct from the speaking hemisphere? This complicated act of cross-cueing, apparently involving observation, memory, volition, inference and foresight, would certainly count as intentional and demonstrative of self-consciousness if a speaking human performed it, say in some game like *charades*. The fact that the mute hemisphere cannot say to itself things like: 'Now I'm going to stick the point of the pencil into my thumb—he's sure to feel pain and if he remembers that only the pencil has a sharp point, he'll guess what it is I'm holding there', does not prove it is unaware of self, but only that it cannot represent self-awareness to itself in a verbal conceptual scheme, at least in so elaborate a fashion.

However, we must not forget that the minor hemisphere does understand simple questions and can respond nonverbally to them in an appropriate manner. The trouble is that it is very difficult to obtain responses from this hemisphere to straightforward questions about its self-awareness because any manifestation of mute hemispheric independence is very upsetting, at least to the speech hemisphere, as was shown in experiments where double and conflicting responses were elicited (Levy, Trevarthen and Sperry [1972]). It is difficult even to test the upper limits of cognitive abilities in the minor hemisphere because of the speaking hemisphere's dominance over the channels of motor expression, its residual inhibiting effect on the minor hemisphere and its constant interference with the left hand's performance, even after surgical disconnection from the right hemisphere (Levy, Nebes and Sperry [1970]).

Accordingly I have suggested to some investigators[1] that if it were medically feasible it should be possible to administer intracarotid amobarbital injections on the left side after experiments like these, temporarily anaesthetising the speech hemisphere, and then ask the mute hemisphere to indicate by head nodding or other non-verbal signs whether the voice heard just before was speaking for *it*. If the answer were clearly negative, one could press on and ask how long it has felt itself to be different from the speaking voice, etc. Depending on the ingenuity of one's protocols, one should be able to get (or fail to get!) fairly direct indications of self-awareness during the two or three minutes the left hemisphere is obtunded without interference from that hemisphere and with little chance that the speaking hemisphere will have any way of knowing what took place. However unilateral amalytisation, while a standard test for speech lateralisation and bilateral memory functions before certain types of surgery (*cf.* Blume, Grabow, Darley and Aronson [1973]) is not without risk (embolism, *etc.*) and can be medically justified only in the presence of clinical indications.

4 THE MINOR HEMISPHERE ON ITS OWN

The circumstance under which the minor hemisphere can best demonstrate its capacities and limitations is also a most tragic one: in the event of non-fatal complete dominant hemispherectomy. Smith ([1966]) has provided us with the best-documented case I know of, a 47-year-old right-handed male who had the whole left hemisphere removed because of a malignant neoplasm. As would be expected the patient, *E.C.*, showed post-operative paralysis of the right side of the body and blindness in the right visual field. There was also severe expressive aphasia and, initially, receptive aphasia as well: indicating clear language dominance in the now missing left hemisphere. However, *E.C.* could enunciate expletives well ('Goddamit!' when he failed to find the words he wanted), and occasionally produced appropriate propositions in response to questions. Nevertheless expressive speech remained severely impaired eight months after the operation.

---

[1] To M. S. Gazzaniga and J. E. Bogen in personal communications in 1973, and more recently to R. W. Sperry and his colleagues in an informal talk I gave at their Psychobiological Laboratory in Caltech, Pasadena on 1 May 1975.

It was in verbal comprehension that *E.C.* made his best recovery. A few months after surgical intervention he correctly selected 85 of 112 items on the Peabody picture vocabulary test. Smith writes:

> On 25 May 1966 I placed five coloured pens (red, blue, green, yellow and black) before *E.C.* and asked him to pick up the yellow pen. When he selected the red one at his extreme left, I carefully explained that this was a test to see if he could recognize colours and he was to pick up only the coloured pen I named. He nodded vigorously, and proceeded to select correctly the randomly named colours on all nine remaining trials (p. 469).

Non-language mental functions continued to show improvement postoperatively too. *E.C.* scored 108 and 104 on the *W.A.I.S.* performance subtests and fifteen years on the Porteus maze test, using his left hand, in May and June 1966. In that same period he also showed he could do simple arithmetical calculations. Of his emotional life, Smith says:

> Affective reactions and general behaviour I have observed before and after hemispherectomy were appropriate, and consistent with his wife's report of no noticeable change in emotional responses or in a basically well-balanced personality. . . . *E.C.* can tell the time, moves about independently in his wheelchair, keeps appointments in the Omaha Veterans Administration Hospital on other floors (requiring use of the automatic elevator) without being reminded, and has been going home at weekends since 26 March 1966. Libidinal drives were reported [by his wife] as normal (p. 470).

Bogen ([1969]) interviewed this patient nearly a year postoperatively. His observations of the speech deficit are worth quoting in full:

> His speech consisted almost entirely of 'Yes, but'—with an occasional 'but-uh, no' (usually at an appropriate time) and from time to time he would say 'God damn it'. On repeated testing, he did not name any object shown to him or put into his left hand. However, when he was shown a pen and was simultaneously told by Dr. Smith, 'say pen', he said 'well, uh-PEN' quite clearly. On one occasion when asked what he wanted to do, he said, 'Uh, duh, uh HOME!' He was then asked, 'Do you mean you want to go home?', and he replied, 'Yes, by God!' When shown a picture of his granddaughter, he smiled; and when asked 'Is this a boy or a girl?', he said, 'girl'. His disability in speaking was in marked contrast to his good comprehension of speech (p. 98).

5 MUTE PERSON OR CONSCIOUS PART OF A PERSON'S BODY?

Now if we are to take Dewitt's analysis at face value, this patient, *E.C.*, is not a person at all. He says:

> The left brain is thus the person I refer to by the name 'John Doe'; there is no more nor no less of John than there ever was. John's right brain now becomes a part of John much as John's arm or leg can be said to be a part of John (albeit in this case a *conscious* part). Previously our naïve tendency was

> to associate John with his whole brain; the splitting of John's brain makes this association untenable. We erred before not so much in associating a person with a brain, but only in associating him with *too much* of a brain (p. 46)!

But surely *E.C. knows* things, like the time of the day, where his appointment is in the hospital, that this is the picture of his granddaughter. Are not entities who can know things of this sort persons? Not according to Dewitt.

> The force of my view would be to deny that the person can be identified with the knower, *tout à fait*. Rather we must recognise that not only can a person know things but a *part* of the person (his minor hemisphere) can also know things. Thus whenever we refer to a given act of knowing it is an open question as to whether or not the person (i.e. the major hemisphere) knows the information in question (p. 46).

Well, then, what does a human being have to have and have to be able to do in order to qualify as a person? Evidently for Dewitt he has to have a left (if that is the speaking one) cerebral hemisphere and be able to talk, write and think propositionally. But of course the two conditions do not amount to the same. Someone with an otherwise healthy major hemisphere can be rendered just as permanently aphasic and agraphic as *E.C.* due to a suitably large lesion of Broca's area at the foot of the third frontal convolution (Brodmann's 44 and 45) on the left side.[1] If full functional language capabilities are the necessary condition of personhood, then there are thousands of twin-brained non-persons sitting around in aphasia clinics throughout the world.

E.C.'s case should be a specially troublesome paradigm for Dewitt. The minor hemisphere survives. It *was*, on Dewitt's reading of the facts, a conscious part of a person. But it no longer *is*, since that person (based in the major hemisphere) is now dead. So a year after the operation we have a conscious entity that looks and acts in many ways like a human being (though hemiparetic, hemianopic and aphasic), a being to which you can talk, at least about simple things, and which shows good verbal comprehension, affect and nonverbal intelligence, but according to Dewitt's analysis is just a surviving, conscious remnant of a former person.

6 THE RELATIVE IMPORTANCE OF PERSONS

At the end of his essay Mr Dewitt acknowledges, quite rightly, that we have no sure guideposts in ordinary language for deciding the numerical aspects of humans. But then he adds these remarks:

> However, the problem is partly semantical and I really care very little whether we are all of us really two 'persons'. I do however intend to insist that the two hemispheres are on radically unequal ground; that the major

---

[1] *E.C.* probably appears to be in far worse shape with a whole hemisphere gone. Actually the same behavioural syndrome could result in someone with left hemisphere present, if to the Broca's lesion we add destruction by infarction of the left calcarine cortex and extensive lesioning of the left sensori-motor cortex. Such a combination of lesions would be rare indeed; since Dewitt does not make visual or motor functions criterial for personhood, it is sufficient to single out the damage responsible for expressive aphasia.

> hemisphere is richer in an important sense than its minor relative. And if we decide that we are all of us really two persons then I shall merely say, 'Fine! But let us just remember that one of us is a much more important person than the other!' (p. 46).

In one sense (but a very selfish sense) my major hemisphere is more important to me than the minor one, since I think it is the base of *my* conscious life (with large inputs from the minor hemisphere, and *vice-versa*) and I, *qua* philosopher, make my living with verbal skills. But it is quite conceivable that another individual, say an artist, would take a different view.

Suppose Picasso was a well-lateralised right hander and that at about age fifty he developed symptoms of a neurological disorder. Suppose this turned out to be a meningioma of the falx. The neurosurgeon might explain to him that in order to get at the tumour from above via the sagittal midline, he will have to retract one of the hemispheres, thereby risking inflammation, swelling and possible damage anywhere on the cortical surface of that hemisphere. Normally there would be no question which half brain to retract in a patient with speech and motor dominance in the left hemisphere. However in Picasso's case the neurosurgeon might explain that there is strong evidence for a greater right hemisphere contribution to artistic activity involving visuospatial perception, such as painting, and that at the risk of losing speech functions he might prefer retraction of the left hemisphere: leaving the choice to him. It is not unreasonable that a man like Picasso (who in real life talked little and wrote almost nothing but painted constantly) might opt for protecting the minor hemisphere.

Now it is a peculiarity of my view that, if Picasso did so and the left hemisphere were badly damaged, the intact survivor would not be the person who took the decision. For this person, unlike the one who told the surgeon he preferred retraction of the left half brain, is congenitally mute. Thus the speaking Picasso in my apocryphal story, by accepting surgical risk to the left hemisphere in order to protect the mute creator based in the right hemisphere, would whether he knew it or not have acted in a spirit of self-sacrifice for the good of mankind.

ROLAND PUCCETTI
Dalhousie University

REFERENCES

BLUME, W. T., GRABOW, J. D., DARLEY, F. L. and ARONSON, A. E. [1973]: 'Intracarotid Amobarbital Test of Language and Memory before Temporal Lobectomy for Seizure Control', *Neurology*, **23**, pp. 812–19.

BOGEN, J. E. [1969]: 'The Other Side of the Brain I: Dysgraphia and Dyscopia Following Cerebral Commissurotomy', *Bulletin of the Los Angeles Neurological Society*, **34**, pp. 73–105.

BOGEN, J. E. [1974]: 'The Dark Side of the Brain', lecture given at Dalhousie University, 28 May 1974.

DEWITT, L. [1975]: 'Consciousness, Mind, and Self: The Implications of the Split-Brain Studies', *British Journal for the Philosophy of Science*, **26**, pp. 41–7.

GALLUP, G. G. Jr. [1970]: 'Chimpanzees Self-Recognition', *Science*, **167**, pp. 86–7.

GLASS, A. V., GAZZANIGA, M. S. and PREMACK, D. [1973]: 'Artificial Language Training in Global Aphasics', *Neuropsychologia*, **11**, pp. 95–104.

HÉCAEN, H. and ANGELERGUES, R. [1962]: 'Agnosia for Faces (Prosopagnosia)', *Archives of Neurology*, **7**, pp. 92–100.

LEVY, J., NEBES, R. D. and SPERRY, R. W. [1970]: 'Expressive Language in the Surgically Separated Minor Hemisphere', *Cortex*, **7**, pp. 49–58.
LEVY, J., TRAVARTHEN, C. and SPERRY, R. W. [1972]: 'Perception of Bilateral Chimeric Figures Following Hemispheric Deconnection', *Brain*, **95**, pp. 61–78.
NEBES, R. D. [1974]: 'Hemispheric Specialization in Commisurotomized Man', *Psychological Bulletin*, **81**, pp. 1–14.
PASSINGHAM, R. E. and ETTLINGER, G. [1974]: 'A Comparison of Cortical Functions in Man and the Other Primates', in *International Review of Neurobiology*, **16** (*ed.* by C. C. Pfeiffer and J. R. Smythies), Academic Press, N.Y., pp. 233–99.
PREMACK, D. [1971]: 'Language in Chimpanzees', *Science*, **172**, pp. 808–22.
PUCCETTI, R. [1973]: 'Brain Bisection and Personal Identity', *British Journal for the Philosophy of Science*, **24**, pp. 339–55.
SAUL, R. and SPERRY, R. W. [1968]: 'Absence of Commissurotomy Symptoms with Agenesis of the Corpus Callosum', *Neurology*, **18**, p. 307.
SMITH, A. [1966]: 'Speech and Other Functions after Left (Dominant) Hemispherectomy', *Journal of Neurology, Neurosurgery, Psychiatry*, **29**, pp. 467–71.
SPERRY, R. W. [1968]: 'Hemisphere Deconnection and Unity in Conscious Awareness', *American Psychologist*, **23**, pp. 723–33.
SPERRY, R. W. [1974]: 'Lateral Specialization in the Surgically Separated Hemispheres', in *The Neurosciences Third Study Program* (*ed.* by F. O. Schmitt and F. G. Worden), MIT Press, Cambridge, Mass., pp. 5–19.
ZAIDEL, E. [1973]: *Linguistic Competence and Related Functions in the Right Hemisphere of Man following Cerebral Commissurotomy and Hemispherectomy*, unpublished Ph.D. Thesis, California Institute of Technology, Pasadena.
ZAIDEL, E. [1974]: 'Language, Dichotic Listening, and the Disconnected Hemispheres', Paper presented at the Conference on Human Brain Function, *U.C.L.A.*, 27 September 1974.

WHAT SORT OF PERSONS ARE HEMISPHERES? ANOTHER LOOK AT 'SPLIT-BRAIN' MAN

1 Studies of the cerebral cortex of man have long since established the proprietary status of the left hemisphere in regard to verbal performance. Instances of language dominance in the right hemisphere are rare even among left-handed people who display left-eye dominance (Penfield and Rasmussen [1950]). This has been strikingly illustrated in recent years by the behaviour of patients who have undergone therapeutic transections of the inter-cerebral commissures (Sperry *et al.* [1969]). These patients can identify palpably but not visually those otherwise familiar objects presented to the left hand. The tentative explanation of their deficit is that the tactile cues delivered to the left hand are ultimately transmitted to sensory centres in the right hemisphere wherein mechanisms associated with speech are either deficient or absent.

In a summoning article on 'split-brain' man, Roland Puccetti has proposed that studies of commissurotomised patients offer evidence of multiple personal identity (Puccetti [1973]). His argument rests upon the putative discovery that such patients are, individually, in possession of two functionally distinct minds. Larry Dewitt has now challenged this by denying that there is '*a self in the minor hemisphere*' (Dewitt [1975]).

I believe that Puccetti has exaggerated the implications of 'split-brain' studies to the issue of personal identity and that Dewitt's attempted rebuttal shares many of the confusions introduced by Puccetti's analysis.

Recall that it is the commissurotomised patient's simultaneous awareness and non-awareness that is supposed to make him relevant to the issue of personal identity. It is of importance, then, to recognise in how many ways this heralded condition can be achieved.

If an observer is presented with a brief flash closely followed by one of greater intensity and duration, he will fail to report the first; this despite the fact that the first is unfailingly reported when no second flash is delivered. The effect is called *backward masking* and it has played an important part in modern theories of visual perception (Raab [1963]; Robinson [1971]). If, under these backward masking conditions, the observer is instructed to press a key as soon as he sees a flash, we can determine the effect the mask has on the processing of the first flash. As it happens, reaction-time to the first flash is unaffected by a presentation of that very masking flash which eliminates *verbal* evidence of detection (Fehrer and Raab [1962]). We might say that the observer has the ability to respond as quickly to a flash he fails to 'see' as to one he does. That is, although he is not aware of the flash he can and does respond to it all the same. But what would it possibly mean to say that this finding is evidence of two *persons* occupying the body of the observer? Yet, Puccetti's argument seems to require just this.

In another class of experiments, a matrix of letters is briefly presented to an observer. The matrix may, for example, be 8×4 or 4×4. With very brief exposures it is found that the normal observer correctly reports between five and nine of the sixteen (or more) items. Loosely, we may say that the span of perception-with-retention is, *ceteris paribus*, about seven items wide. On any given presentation, there will be some ten items apparently unperceived. Suppose, however, that immediately after the display disappears, a cueing ring is projected over one of the matrix locations. We now find that the observer can tell us the letter that previously occupied this location even though, on the trial in which the letter occurred, he failed to report it (Sperling [1963]). Here again the observer simultaneously 'knows' and does not 'know' that the letter in question was part of the array. Without the cue, he will swear that he has reported everything he has seen. With the cue his span of retention is widened, *ex post facto*. If trials without cues correspond to our observer in state-A and those with cues to him in state-B, we must conclude that he does not know in A what he reliably discloses in B. On Puccetti's account, two minds and, therefore, two persons are operating.

There are exposure durations so brief that an attentive subject cannot identify an intensely illuminated word. However, if it is presented repeatedly—with the interval between successive presentations sufficiently long to prevent visual persistence—the subject will come to identify the word reliably. That his success is not merely the result of piecemeal additions of partial information is proved by the fact that the effect occurs even with Turkish words and other items foreign to the observer's vocabulary (Haber and Hershensen [1965]). Note the difficulty in accounting for the observer's performance: on no single previous trial was he successful in reporting the word but on *this* trial he is. If his performance is to serve as a reflection of what is in his mind, there are two minds; one possessing this strange Turkish word and one insisting that the word was never seen. Two persons?

Even the paradoxical ability of a patient to recognise familiar objects held in the hand but not to recognise these same objects when presented for visual inspection is not without precedent. Visual *agnosias* of this sort are common in the neurological literature and there are auditory and tactile agnosias as well (Bannister [1973]). The same patient who names the touched object correctly but cannot name the same object when it is placed before him gives us one more instance of simultaneous 'knowing' and 'not-knowing' and this despite the fact that he is normal with respect to visual sensitivity. These agnosias are nearly invariably associated with left-hemisphere disturbances and they form a genre of which the commissurotomy symptoms are an instance.

The experimental and clinical findings reviewed above could be prolifically multiplied. What they have in common is their reflection of the nuances of human information-processing. Observers will respond as quickly and consistently to 'unseen' flashes as to those they report. Unrecalled items will quickly become restored to memory by a cueing marker which is otherwise devoid of informative content and which has nothing in common, except location, with the 'unknown' item. Observers will suddenly report strange words that were consistently 'unseen' on all previous presentations. Patients will simultaneously possess and not possess objective knowledge. We have dreams during which we know we are dreaming. We might even dream of knowing we are dreaming. Psychiatry has unearthed a rich trove of repressed material; memories that somehow are and are not accessible to the same patient. And there is the renowned mnemonist whose ability to perspire on request merely by thinking of himself running after a bus has been so graphically described by A. R. Luria ([1968]). In all these instances the percipient judges himself to be one person. There are, of course, cases of multiple personality (*e.g.*, the 'three faces of Eve') and instances of amnesia in which there is not the *simultaneous* operation of what Puccetti calls 'two minds'. Nonetheless, even these cannot be ruled out of Puccetti's argument since they provide evidence of at least two sets of memories that can be evoked separately by therapeutic means. In fact, the more one considers the range of psychiatric and neurological disturbances, the more one reflects on the oddities of day to day perceptions and recollections, the more one is constrained to believe that an instance of single identity cannot be found. Quite simply, the findings resulting from studies of commissurotomised patients are not *sui generis* in the sense required by Puccetti's thesis.

2 If the question is whether the findings and observations reviewed above offer evidence of multiple minds, the answer is *yes* provided that what we mean by this is no more than the parallel processing of information in such a way as to produce conflicting and even contradictory outcomes in retrieval tasks. What the observer reports is not always a reliable measure of what he knows or has perceived. This is so for the unremarkable reason that acquisition, storage, and retrieval are different processes and, accordingly, a single, dimensionless (YES–NO) response cannot disclose the status of all three. (It is principally because of this that Locke's attempt to reduce personal identity to memory had to fail.)

Puccetti's analysis of personal identity is not of the orthodox Humean stripe but it does court the sorts of difficulties that infect Hume's treatment of the issue. Chief among these is what might be called *the fallacy of the man in the mirror*. It is this fallacy that causes the untutored to wonder why, since left and right are reversed by a mirror, top and bottom are not! The fallacy is committed by an observer who treats his reflected image as if it were someone else and who thereupon assumes that this other person's left hand is to the observer's right. Hume, in his discussion of personal identity (*Treatise of Human Nature,* I, IV, VI), rejects the 'metaphysical' assertion of a self because, when he looks for that self, he cannot find him. Where the fallacy enters into Hume's account is, of course, at the point at which he makes a distinction between the 'he' doing the looking and the 'he' who is sought. It is the looking, not the finding, that confirms the self.

A similar form of the fallacy is provided by Puccetti whose treatment of personal identity is akin to the Humean '*bundle of sensations*'. Since the commissurotomised patient reveals or at least reports contradictory sensations, since he is unable to recall occurrences which he has so recently responded to, two minds must be at work, according to Puccetti. Somehow Smith is different; he no longer names items and he forgets ever having seen them when they have been delivered to his right hemisphere. Alas, there is a new and additional Smith.

But as we have shown, the non-commissurotomised Smith will yield the very same results in a variety of contexts. Indeed, one way of identifying Smith, at least with respect to *genus* and *species*, is that he is the sort of organism (*a*) whose reaction-time to masked flashes is no slower than that to unmasked flashes; (*b*) whose percepts grow in a certain way; (*c*) whose memory can be cued after the fact such that he now recalls that which he could not recall. Lastly, Smith is also the fellow who perceives and behaves in a predictable way if he happens to have his *corpus callosum* split.

The reader may be ready to note that these characteristics are scarcely sufficient to identify Smith since they are shared, or so we have every right to expect, by all human beings tested in the appropriate ways. And this is just the point: perceptions and behaviour following commissurotomy are not affected uniquely for Smith. He is as surprised as the victim of 'phantom limb' who awakens to discover that his painful appendage was left on the battle field. He is as surprised to learn what is going on with his performance as he is when he sees some other commissurotomised patient. Were there two Smiths, one of them would not be surprised.

I have limited my analysis to that portion of Puccetti's thesis which rests upon studies of "split-brain" man. However, Puccetti also raises the issue of multiple identity in the very different context of transplanted hemispheres. Here, one corpse receives Smith's right hemisphere and another corpse receives his left. We are now asked to welcome two Smiths since these vivified bodies are now in possession of the same memories, perceptions, attitudes, etc.

The problem with this facet of Puccetti's thesis is that it is either tautologous or wrong. If, *by definition*, we must accept that everything comprising "personality" is a brain process and that each hemisphere mirrors the contents and activities of the other, then, *necessarily*, two bodies each possessing one of Smith's hemispheres are Smith. But personality, perception, attitudes and the like are not "things". Nor is the cerebral cortex a passive repository of experience. The functional (and, to a certain extent, structural) characteristics of the central nervous system change with experience, with age, with disease, with diet. In this sense, each person comes to possess a great number of "identities" and, to this extent, Socrates seated and Socrates standing may not be the same "person".

Since brain function is not a constant over the life of the individual and since we already know that the two hemispheres are neither symmetrical in function nor identical in, to use a less than felicitous term, content, we have no reason to expect that transplanted hemispheres will constitute transplanted identities.

Once it becomes inviting to admit divided attention, cued retrieval, perceptual incongruities, etc. as evidence of multiple minds, any number of metaphorical uses of the nervous system are possible. Thus, Dewitt will rescue single identity from Puccetti's thesis by insisting that one of the hemispheres is not a mind at all: '. . . The minor hemisphere is deficient in just this respect; that it lacks the very awareness of self which is so integral in making the human mind what it is (Dewitt [1975], p. 43). And a little later, '. . . the absence of self in the minor hemisphere . . . is sufficient to disqualify it as a person' (p. 45).

The plain fact is that hemispheres are not predicated of failures, confusions, contradictions, percepts, judgments. Need we remind ourselves again that a wedding invitation filed in the middle drawer does not confer an invitation on the drawer? We have every right to describe post-operative Smith as feeling restored after a walk in the garden and a spinach soufflé but, short of humour, we would not discuss hemispheres, minor or major, this way.

Nor is it beside the point to mention that even the clinical data do not conform to a single or simple pattern. The commissurotomised patient typically has had a long, severe and tragic history of repetitive seizures leaving his brain scarred, atrophied, and locally dysfunctional. Added to this is the radical nature of the surgery itself. The verbal reports of such patients, while constituting the only relevant information about what is 'in' their minds, are not reliable evidence. It is not the kind of evidence that is safely generalised to the otherwise normal population. The most consistent finding of psychological consequence remains their inability to name objects held in the left hand. However, they can *point* to this very same object to indicate that it was the one they recently held. How is this possible without some form of symbolic mediation? The fact is, it is impossible. Berkeley observed long ago that an object does not 'feel' round the way it 'looks' round; that 'round' is the word by which otherwise different experiences come to be treated as if they were the same. The blind-folded patient's ability to point at what he held is evidence enough that some symbolic processes are mediated by the non-dominant hemisphere. Are these processes linguistic? What sort of persons are hemispheres? Such questions are so rife with ambiguity as to make replies impossible or implausible. If recent research with chimpanzees has convinced some that the chimps are engaged in language, it should be recalled that these animals do not display cerebral dominance in the required sense (Pribram [1971]).

Finally, that these patients will be called upon to serve more than the problem of personal identity is already becoming sadly apparent. In a tolerant citation, Dewitt refers to Robert Ornstein's peculiar simile: the left hemisphere, like Western philosophy, is analytical, verbal, linear, and rational while the right, like Eastern thought, is intuitive, non-verbal, non-linear, a-rational (Ornstein [1972]). (The inference, I suppose, is that Orientals tend to be left-handed.)

Before this loose talk about hemispheres, minds, selves, and culture becomes habitual we are advised to economise in the use of the human nervous system for metaphorical purposes. The lateralisation of speech mechanisms has been a secure fact in neurology for more than a century. Disturbances of language and perception come in great variety with no two very alike. Some are expressive, some receptive, some mixed. Some are irreversible, others remediable. The consequence of treating these diseases as evidence of new persons will be to render the concept of person inarticulate and that of pathology virtuous.

DANIEL N. ROBINSON
Georgetown University

REFERENCES

BANNISTER, R. [1973]: *Brain's Clinical Neurology.*
DEWITT, LARRY W. [1975]: 'Consciousness, Mind, and the Self: The Implications of the Split-Brain Studies', *British Journal for the Philosophy of Science*, **26**, pp. 41–6.
FEHRER, E. and RAAB, D. [1962]: 'Reaction Time to Stimuli Masked by Metacontrast', *Journal of Experimental Psychology*, **63**, pp. 143–7.
HABER, R. N. and HERSHENSEN, M. [1965]: 'The Effect of Repeated Brief Exposures on the Growth of a Percept', *Journal of Experimental Psychology*, **69**, pp. 40–6.
LURIA, A. R. [1968]: *The Mind of a Mnemonist.*
ORNSTEIN, R. [1972]: *The Psychology of Consciousness.*
PENFIELD, W. and RASMUSSEN, T. [1950]: *The Cerebral Cortex of Man.*
PRIBRAM, KARL [1971]: *Languages of the Brain.*
PUCCETTI, R. [1973]: 'Brain Bisection and Personal Identity', *British Journal for the Philosophy of Science*, **24**, pp. 339–55.
RAAB, D. [1963]: 'Backward Masking', *Psychological Bulletin*, 60, pp. 118.
ROBINSON, D. N. [1971]: 'Backward Masking, Disinhibition, and Hypothesized Neural Networks', *Perception and Psychophysics*, **10**, pp. 33–5.
SPERLING, G. [1963]: 'A Model for Visual Memory Tasks', *Human Factors*, **5**, pp. 19–31.
SPERRY, R. W., GAZZANIGA, M. and BOGEN, J. E. [1969]: 'Interhemispheric Relationships: the Neocortical Commissures: Symptoms of Hemisphere Deconnection'. In P. J. Vinken and G. W. Bruyn (*eds.*): *Handbook of Clinical Neurology*, **4**, pp. 273–90.

Brit. J. Phil. Sci. 27 (1976), 79–92 *Printed in Great Britain*

Review Article

WAS FREGE A LINGUISTIC PHILOSOPHER?*

> 'Frege can thus be considered the father of "linguistic philosophy" . . . '
> Michael Dummett, *Frege. The Philosophy of Language*

> ' . . . I am compelled to occupy myself with language, although it is not my proper concern here'
> Gottlob Frege, *Der Gedanke*

1 The story of how Frege was ignored by most of his contemporaries, and misunderstood by those who troubled to notice his work is well known.[1] Today Frege is one of the most discussed figures in the history of philosophy, but even so it is far from clear that his work has been properly understood. Many of the most notable commentators on Frege's philosophy have been writers deeply influenced by Wittgenstein, and there has been a consequent tendency to see Frege's philosophy in the light of Wittgensteinian problems. This has, I think, resulted in the currency of a false impression of Frege's philosophical achievement, and nowhere is this false impression more systematically conveyed than in Michael Dummett's recent and much heralded volume on Frege and the philosophy of language. First then, a word about the philosophical background to this book.

2 Dummett places himself in a class of philosophers for whom philosophy starts with an analysis of meaning—'linguistic philosophers' as he calls them.[2] He obviously finds much inspiration in the work of Wittgenstein, but he explicitly rejects the 'philosophy of ordinary language', because its uncritical attitude towards usage lead to a naive view of language and to the 'notorious "paradigm case argument" '.[3] Further, Dummett is willing to incorporate into his analysis the methods and results of mathematical logic—a subject that Wittgenstein despised.[4] But Dummett does accept a central Wittgensteinian idea: the theory of language games,[5] and this idea leads to some of the 'ordinary language' philosophy Dummett rejects, and to the rejection of the kind of universal logic which he accepts. A main purpose of Wittgenstein's doctrine of language games was to delimit the sphere of legitimate linguistic activity: a

---

* Review of M. Dummett [1973]: *Frege. The Philosophy of Language*. London: Duckworth. £10. Pp. xxv+698. I would like to thank Professor Irving Copi and the members of the Department of Philosophy, London School of Economics, especially Colin Howson for their careful criticism of earlier drafts of this paper.

[1] Husserl, Russell and Wittgenstein were exceptions to this, but their interest came after the most creative period of Frege's career.

[2] Dummett [1973], p. 683.

[3] *Ibid.*, p. 455.

[4] *Cf.* Wittgenstein [1956], iv, 48.

[5] Dummett *ibid.*, *e.g.* pp. 3, 34, 310, 331.

particular usage can be justified if and only if it is embedded in a language game.[1] Language games *insulate* usage from criticism. Further, each such game is autonomous. There can be no *universal* standard of validity, for such a standard would make language games comparable.[2] It follows that mathematical logic is illegitimate, since it attempts to impose, from outside, certain artificial standards on the mathematicians' language game.[3] So Dummett cannot easily jettison those parts of Wittgenstein's philosophy which do not suit him, and his position, which sounds like an improvement on old style Oxford philosophy, cannot be taken simply at face value.[4] At this point the reader's difficulties with Dummett's book begin.

3 Since the book nowhere contains a really clear statement of the author's theses, or of his general philosophical point of view, it is extremely difficult to resolve the sort of question that I have just raised. Dummett, it appears, does not want to linger over preliminaries, but instead is eager to confront each separate topic. Against this, I think that in a book of seven hundred pages, a clearer statement of motivation and intention is essential if the reader is adequately to assess its merits. As it is, each lengthy chapter meanders through a variety of positions. Theses are stated, objections raised, alternatives proposed; this process being all the while broken up by the interjection of new ideas and vague references to earlier stages. Rarely is a position clearly stated and defended. As I shall try to show,[5] the central thesis of the book (concerning the proper approach to philosophy) is scarcely argued for at all, though it is often appealed to as if it were a piece of sound common sense.

A connected difficulty is the prolixity of the book. The following example will illustrate.

> It [Frege's theory of sentence construction] allows the use only of the simple account of the truth-conditions of sentences containing signs of generality that is adequate for those involving only simple generality, as a general account applicable to all signs of generality, provided that the application is made only to the stage of construction at which the sign of generality in question is introduced.[6]

A comprehensive subject index would have been invaluable, but Dummett merely offers a list of sixteen, rather idiocyncratically chosen topics. This is justified on the rather peculiar ground that anyone interested in Dummett's views on one subject will have to read the whole book anyway.[7]

When it comes to Dummett's treatment of Frege's own views, the reader is once again in difficulty. The book is almost entirely free both of quotations from Frege, and of specific references to the relevant passages in Frege's work. This practice—and Dummett is not the first to follow it[8]—is always deplorable, and particularly so when factual errors creep into the expositor's account. For instance, Dummett sets himself the task of explaining why Frege held that we can make inferences only from true premisses, rather than from *asserted premisses*, the latter being more consonant with the rest of his logical theory.[1] But, as has been shown elsewhere,[2] Frege's formulation of this point was at least ambiguous. Indeed, when seen in their context, statements like 'Only true thoughts can be premises of inferences',[3] can be seen as a careless shorthand for 'only thoughts taken to be true (*i.e.* asserted) can be premisses of inferences'. If correct, this explodes Dummett's attempt to solve a problem that does not exist. But correct or not, it highlights the deficiency of a book which does not give the reader an opportunity to decide for himself.[4]

---

[1] *Cf.* Wittgenstein [1951], pp. 22, 137.

[2] Wittgenstein [1951], p. 224. According to Wittgenstein, language games are *related*, but the relations between them cannot be sufficiently articulated to be given formal, logical expression (*cf.* Wittgenstein [1951], paragraph 66).

[3] Wittgenstein [1956], v, 13.

[4] See *below*, p. 90, text to footnote 1.

[5] See *below*, section 4.

[6] Dummett *ibid.*, p. 11.

[7] *Ibid.*, p. 695.

[8] Geach [1961] also contains no references.

The rather tortuous style of this book is, at least in part, a consequence of Dummett's decision to relate Frege's own work to subsequent developments in philosophy, many of which took a very different direction from anything envisaged by Frege. Such an approach can be worthwhile if the aim is to shed light on subsequent problems by reference to solutions implicit in Frege's work. But Dummett's purpose is rather the reverse. He wants to *appraise* Frege's philosophy by relating it to later developments.

> The book attempts, not only exposition, but also an evaluation . . . It is impossible adequately to evaluate Frege's doctrines without forming opinions about topics of which he did not treat, . . . Some sections of the book are therefore hardly about Frege at all, but about matters which must be considered if one is to judge whether Frege spoke the truth.[5]

This is certainly a startling claim. Surely, in order to decide whether or not somebody has spoken the truth we need not look further than what he said, and the logical consequences thereof.[6]

Proceeding in this way, Dummett presents and contrasts the views of Frege, Russell, Wittgenstein, Quine, the author himself, and other contemporary philosophers on a variety of topics. The result of this collapsing of historical perspective is twofold. First, the task of disentangling the variety of positions is made even more difficult. But more importantly, because the emphasis is on later developments rather than earlier ones, it becomes almost impossible to get a clear view of *Frege's* achievement, which the quotation above suggests was one of Dummett's aims. Frege's achievement was, quite simply, that he made startling progress on logical and philosophical problems that had proved intractable at the hands of his predecessors and contemporaries. This fact is not one whit affected by subsequent developments, however interesting they may be in themselves. One can measure Frege's greatness only by seeing his work in the light of his problem situation and the research traditions which had developed prior to his work. But this book contains only passing references to Boole, Cantor, Schröder, and other contemporary researchers in logic. Bolzano is not mentioned at all.[1]

---

[1] Dummett *ibid.*, p. 313.

[2] Stoothoff [1963].

[3] Frege [1906], p. 319. (Page references to this and other articles are to the reprints in *Kleine Schriften.*)

[4] An oversight of a rather different kind occurs on the next page, where Dummett says 'I know no passage where he [Frege] excludes the possibility of asserting what is in fact false' (p. 314). As a statement about Dummett this may well be true, but if we replace 'I know' by 'There is', the claim seems to be refuted by a passage from *Die Verneinung* where Frege says of the negation of a true sentence, 'since it is false, it may not be uttered assertively (*behauptender Kraft*)' (Frege [1918*b*], p. 367).

[5] Dummett, *ibid.*, p. xii.

[6] Certainly Dummett's proposal would be very damaging to the fairness of judicial proceedings!

4[2] Dummett's (neo-Wittgensteinian) view of philosophy and his interpretation of Frege are most clearly expressed in the last chapter: 'Frege's Place in the History of Philosophy'. Here Dummett makes two claims: The first is that the analysis of meaning is the central problem of philosophy. The second claim is that the recognition of this priority is Frege's greatest contribution to philosophy.

According to Dummett, 'philosophy has, as its first if not its only task, the analysis of meanings'.[3] He does not claim—as others have before him—that this analysis makes subsequent philosophising redundant by revealing the problems of philosophy as pseudoproblems. Meaning analysis is simply *prior* to the application of philosophy to particular problems.

> . . . until we have first achieved a satisfactory analysis of the meanings of the relevant expressions, we cannot so much as raise questions of justification and truth, since we remain unclear about what we are attempting to justify or what it is about whose truth we are enquiring.[4]

But it turns out that Dummett is not so much interested in meanings, but in Meaning (with a capital 'M'). He admits that we often know what we are talking about without having an explicit theory of meaning. But he claims that this is not the case with certain key philosophical terms. In these cases our choice of a *theory* of meaning is crucial.[5]

In favour of this last thesis, Dummett seems to argue that there is a clear distinction within natural language between words which are of philosophical interest and those which are not. For the latter we are easily able to give adequate definitions, in which case the problem of meaning-in-general does not arise, but in the case of the former we cannot. In these cloudy cases we need to invoke a theory of meaning.[6]

These arguments are sketchy and it is not difficult to find objections to them. In fact the position (that meaning analysis is primary) is presented by Dummett in so perfunctory a way as to give the impression that he is presenting a thesis that does not require any strenuous defence. One of the surprising things about this book is that it contains so little in the way of systematic argument defending the importance of meaning. Some of the book's remarks on this subject seem quite ill-considered; for instance the suggestion that the problem of meaning is philosophically important because it is difficult.

---

[1] There are some very brief but perceptive remarks about the relation of Frege's quantification theory to the earlier theory of *suppositio* (see pp. 10–14).

[2] While writing an earlier version of this review, I received a paper by Hans Sluga, entitled 'Frege as a Rationalist'. This anticipates some of the points made here about Dummett's interpretation of Frege, though we discuss this problem in quite different contexts. I should like to direct the reader's attention to this interesting paper, which will appear in M. Schirn (*ed.*): *Studies in Frege*, I, Stuttgart, 1976.

[3] Dummett *ibid.*, p. 669.

[4] *Ibid.*, p. 667.

[5] *Ibid.*, pp. 668–9.

[6] *Ibid.*, p. 669.

> Only when it is clearly grasped how hard it can be to attain an adequate analysis of the meaning of an expression . . . can the analysis of meaning be seen as a primary task of philosophy.[1]

But if difficulty is taken to be the essence of the question, then, for instance, learning Chinese ought to be regarded as a serious philosophical problem.

5 This briefly, is Dummett's view of philosophy, and it is a philosophy which he claims to share, in its broad principles, with Frege. According to Dummett, Frege was a great philosopher, on the same level as Descartes and Kant, because he showed the theory of meaning to be the foundation of all philosophy.

> For Frege the first task, in any philosophical enquiry, is the analysis of meanings,[2]
>
> . . . because, the deeper such analysis goes, the more it is dependent upon a correct general account of meaning, a model for what the understanding of an expression consists in, the theory of meaning, which is the search for such a model, is the foundation of all philosophy, and not epistemology as Descartes misled us into believing. Frege's greatness consists in the first place, in his having perceived this . . . We can, therefore, date a whole epoch in philosophy as beginning with the work of Frege, just as we can do with Descartes.[3]

Dummett correctly points out that Frege's interest in language goes beyond the solution of problems required purely for the purposes of mathematical logic.[4] Where he is mistaken is in contending that Frege's interests in the wider aspects of language, particularly the notion of sense, show that 'he wanted to give a general account of the workings of language, . . . '.[5] Frege's interest in language was highly selective, and his principle of selection was based on prior epistemological considerations. To briefly illustrate this point I shall concentrate sharply on one central aspect of the problem, Frege's theory of sense. This theory was, I claim, designed to solve an epistemological problem, and the form that the theory took was dictated by epistemological considerations.

Frege's problem was how to distinguish statements of identity which differ as to their 'cognitive value' (*Erkenntniswerte*).[6] Statements like 'the morning star is the same as the evening star' 'contain very valuable extensions of our knowledge', while statements like 'the morning star is the same as the morning star' do not. The difference between these sentences concerns the knowledge they express, and particularly the fact that the latter appears to express no (non-trivial) knowledge at all. Frege's discussion of this subject concludes with the remark that '. . . from the point of view of knowledge, the sense of a sentence, namely the thought expressed by it, is no less relevant that its reference, . . . '.[1]

---

[1] *Ibid.*, p. 667. [2] *Ibid.*, p. 667. [3] *Ibid.*, p. 669.

[4] *Ibid.*, pp. 83 and 90. The T.L.S. reviewer missed this point. According to him, in order just to show that the rules of inference preserve truth Frege had to give 'a deep account of what in general, a theory of meaning for a language should be like'! (*Times Literary Supplement*, 30 November 1973).

[5] *Ibid.*, p. 83. [6] Frege [1892], p. 143.

There are two objections which might be raised against my thus construing Frege's problem situation:

(1) Knowledge is something possessed by persons rather than by propositions, and it is wrong to characterise a difference which may arise among propositions as an epistemological difference.

If this argument were correct, the problem would remain as to why Frege expressed himself in terms of knowledge rather than in terms of meaning (see the quotations directly above). But it is not necessary to accept this argument, for there is a clear, well-accepted sense of knowledge according to which it is a property of propositions rather than of the human psyche; as when the question is raised (as it was by all philosophers from Locke to Kant) 'does Newtonian Mechanics constitute genuine knowledge?' This question does not concern the beliefs of individuals, but the epistemological status of a theory.[2] As I shall try to show below, Frege was the first person to clearly recognise the distinction between the subjective and objective aspects of knowledge, and to confront the problems raised thereby.

(2) Whatever Frege said or meant, it is open to Dummett (or anybody else) to *reinterpret* Frege's problem as a problem about meaning, since a difference in 'cognitive value' between two sentences is also a difference in their meaning.

This is weaker than the thesis that Dummett actually holds, *viz.*, that Frege actually regarded the theory of meaning as primary, but it is none the less tenable. Considerations which tell decisively against *this* position only appear when we examine the arguments that Frege elaborates in favour of his theory, since these cannot be reinterpreted as arguments about meaning. To these I now turn.

Frege defines the sense of a complete, declarative expression as a 'Thought' (*Gedanke*); what we would call a proposition. A Thought, as Frege presents it, is essentially epistemological; the 'cognitive content' of an expression,[3] and his arguments in favour of his notion of sense are epistemological ones. They attempt to show that the view that sense is something essentially psychological leads to an intolerable relativism: the impossibility of interpersonal standards of appraisal. If the psychologistic view were accepted, Frege says:

---

[1] *Ibid.*, p. 162. Dummett seems to acknowledge this point when he says 'Frege later took as the starting-point for the sense/reference distinction, how true identity-statements can be informative, . . . ' (Dummett [1973], p. 630).

It may be worthwhile to note that Dummett does not regard reference as a part to Freges theory of meaning. But his reasons for this are decidedly un-Fregean, *viz.*, that 'a theory of meaning is a theory of *understanding*' (p. 92). Since, the argument continues, one can understand ('grasp') the meaning of a word or expression without knowing its reference, the concept of meaning does not directly involve considerations of reference (though sense does 'determine' reference). Whatever the merits of this argument as it concerns 'understanding', it falls down as an explication of Frege's theory of meaning, because meaning for him is something *objective*, and quite unrelated to considerations of understanding. For example, Frege would no doubt have defended the possibility that an expression so complex as to be beyond human comprehension, might still be meaningful (see *below* p. 90).

[2] See Popper [1972], chapter 3.

[3] 'I understand by Thoughts, not the subjective activity of thinking, but their objective content . . . ' (Frege [1892], p. 148, note).

> . . . I, perhaps, have my own science, namely a collection of thoughts whose bearer I am, another person has his science. Each of us occupies himself with the contents of his own consiousness. A contradiction between the two sciences would be impossible and it would be futile to dispute about truth, . . . [1]

But if we are to have an intersubjective science, 'a science common to many, and on which many can work',[2] then we must recognise that this science must consist of objective propositions, not of subjective ideas or beliefs.

This argument recurs in similar contexts throughout Frege's work. For instance in the introduction to the *Grundgesetze* he points out that, if the laws of logic were taken to be psychological laws of thinking, they would be robbed of any ability to arbitrate between conflicting arguments. We could only register a disagreement and say: these laws hold for one person, others hold for somebody else. Clearly rational argument cannot proceed beyond this point, so psychologism can only lead to an increased dependence on non-rational considerations.[3]

Now Dummett actually interprets this attack on psychologism as part of a polemic against incorrect theories of meaning.

> When Frege engages in polemic against psychologism, what he is concerned to repudiate is the invasion of the theory of meaning by notions concerned with mental processes, mental images, and the like, and the confusion between the process by which we come to aquire a grasp of sense, and what constitutes such a grasp.[4]

But, as I have tried to show, the arguments that Frege uses in defence of his own objective notion of sense and against 'mental imagism', as we might call it, are not concerned directly with meaning. In particular they are not concerned with 'what constitutes a grasp of meaning', as Dummett seems to be suggesting in the quotation above and elsewhere. On the contrary, Frege is constantly at pains to point out that the sense of an expression is something that exists independently of anyone's having in fact grasped it, and he would have regarded the admission into his theory of questions about our (subjective) grasp of sense as an intrusion of psychologism.[5]

(Dummett refers to an example in Frege's work which seems at first sight to support his interpretation and contradict my claim directly above.[6] In the *Grundlagen* Frege considers the suggestion that 'parallel' can be defined in terms of 'direction'. But he dismisses this suggestion because to do so would be to invert the 'true order of things'. We intuit the concept 'parallel' and then proceed to the concept of direction *via* an intellectual construction (Frege [1884], p. 75). So Frege seems to be suggesting that a theory that assigns meanings to terms should take into account the order in which meanings are actually grasped. But this example is misleading, and derives from Frege's peculiar views on geometry, not from his views about meaning. Frege made a sharp distinction between geometry, which is based on intuition, and arithmetic, which is to be justified on purely logical grounds. In the case of geometry alone, the (logical) order of justifications corresponds to the (psychological) order of knowing, and the latter is a criterion of the former. Given this premiss, we can reinterpret Frege's argument about the priority of 'parallel' over 'direction' as follows. (*a*) In geometry axioms and definitions need to be justified by an appeal to intuition. Therefore (*b*) that which is psychologically prior is also logically prior. Therefore (*c*) in geometry concepts may be defined only in terms of other concepts which are psychologically prior. In other words, Frege is defending the priority of 'parallel' over 'direction', not *because* we grasp the concepts in this order, but because the fact that we do so *implies* that the one is logically prior to the other. It is always logical priority which counts, but in the case of geometry alone, the psychological order reflects the logical order. Dummett is wrong to construe a particular example as a general thesis.[1])

---

[1] Frege [1918*a*], p. 353.

[2] *Ibid.*

[3] Frege [1893], pp. xv–xviii. My argument against Dummett's own 'descriptivism' is a direct adaptation of Frege's argument here (see *below*, pp. 89–90).

[4] Dummett *ibid.*, p. 240.

[5] It may be replied that a 'grasp' of meaning is to be judged in terms of behaviour, and that this avoids psychologism. But such a position suffers from the same defect: it turns philosophy from a normative subject into a descriptive one. Instead of reducing philosophy to speculative psychology, it reduces it to descriptive sociology (see Lakatos [1976*b*]).

[6] *Ibid.*, pp. 240 and 678.

Dummett's interpretation of Frege's philosophy runs into a further difficulty. I think that any general picture of Frege's philosophy should try to substantiate the impression that his various doctrines have an 'organic unity' which makes them in some sense all part of the same philosophy. As it stands, this is terribly vague, but Dummett agrees with something like it: 'More than that of most philosophical writers, Frege's work can be treated as a unity'.[2] This statement aside, I think that Dummett's whole approach precludes a unified interpretation, because it cuts off Frege's general philosophy from his philosophy of mathematics.

This is because Frege's philosophy of mathematics is quite clearly an epistemological enterprise, closely related to his attack on relativism. His aim was to show that arithmetic and analysis can be justified by their completely rigorous deduction from axioms of pure logic, where deduction is to be characterised in purely formal terms, without the necessity of any appeal to intuition.[3] The following is a characteristic statement of the philosophical convictions behind this programme.

---

[1] I do not claim that this argument of Frege's is a good one: in fact it exhibits the kind of 'logical essentialism' which so much of Frege's work suffers from. The theory that logic picks out a unique order of definition and of deduction is implicit in Frege's writings, and his remarks on 'indefinables' depend upon it (see, *e.g.* Frege [1914], p. 254). He would have regarded the possibility of interchanging defined and undefined terms; axioms and theorems, as an arbitrariness to be avoided by a properly constituted logic.

[2] Dummett, *ibid.*, p. 628.

[3] Dummett calls Frege's axiomatic approach to logic 'retrograde', and he advocates the used of Gentzen-type natural deduction systems instead (p. 432). But given Frege's problem situation and his (forgivable) ignorance of developments that took place after his death, it is easy to see why he favoured the axiomatic approach: he wanted to specify exactly the assumptions that are required in order to derive the whole of mathematics. Dummett seems to ignore this point. He actually criticises Frege's characterisation of logic as the science of 'the laws of truth' because 'In philosophy it led to a concentration on logical truth and its generalisation, analytic truth . . . ' (p. 433). Now regardless of

> Because there are no gaps in the chains of inference [of the *Grundgesetze*] every 'axiom' . . . upon which a proof is based is brought to light; and in this way we gain a basis up on which to judge the epistemological nature of the law that is proved . . . Only if these transitions are split up into logically simpler steps can we be persuaded that the root of the matter is logic alone.[1]

Clearly the formalisation of mathematical proof was a means towards the achievement of an epistemological goal—the discovery of the assumptions upon which the theorems are based, by which we can gain an insight into their justification.

Now if it be admitted the Frege's philosophy of mathematics had the essentially epistemological aim of providing a foundation for mathematics, then I think Dummett will find it difficult to make Frege's philosophy of mathematics cohere with his general philosophy (at least as Dummett interprets it). For, according to Dummett, Frege's achievement in general philosophy was to show that epistemological considerations are of secondary importance.[2]

In fact the book does contain an implicit answer to this problem, even though the problem itself is never acknowledged. It turns out that Dummett regards Frege's philosophy of mathematics as of little intrinsic value, exactly because it is not formulated in terms of problems about meaning.

> . . . his philosophy of mathematics is a starting point only in a historical sense. It is not true that Frege's formulation of the problems of the philosophy of mathematics any longer strikes us as the most fruitful way to pose the questions that arise in that subject . . . [3].

Dummett's reason for this is that '. . . Frege, as a philosopher of mathematics, is indisputably archaic in a sense in which Hilbert and Brouwer are not'. He attributes this to the fact that these later philosophers concentrated on the problem of meaning, almost to the exclusion of the problem of justification: '. . . it was tacitly assumed that, once the meaning of a statement is given, the nature of its truth (if true) would look after itself'.[4] This, however, is quite wrong. Certainly both Hilbert and Brouwer regarded parts of classical mathematics as meaningless,[5] but their reason for doing so was exactly because the admission of the infinite seemed to them to destroy the possibility of justifying the truth of mathematical statements. Like Frege, their central interest was an epistemological one, and their criteria of meaning were designed to solve the epistemological problem: how is mathematics to be justified? If there is an

---

whether or not these later developments were retrograde, this is hardly a fair criticism of Frege. One is not driven to an interest in 'analytic' truth as a *consequence* of holding that the business of logic is the generation of logical truths. Dummett, it seems, wishes to visit the sins of the children on the father! This is a further example of Dummett's conflation of philosophical and historical problems (see *above*, pp. 81–2).

[1] Frege [1893], p. vii. See also his [1879], p. 5.

[2] See the quotations to this effect *above*, p. 83.

[3] Dummett [1973], p. xvi.

[4] *Ibid.*, p. xx.

[5] Hilbert regarded infinitistic statements as neither true nor false. They were only to be employed if their use could be justified by a finitistic consistency proof.

archaism in Frege's philosophy of mathematics it is one that he shares with later philosophers like Hilbert and Brouwer: the idea that mathematics can be justified in some ultimate sense.[1]

6 I will now comment briefly on Dummett's general philosophical position, as I have outlined it in section 4 above. First, let me say that in attacking Dummett's view that philosophy consists of a hierarchy of disciplines with the theory of meaning as ground level, I do not wish to substitute a *different* ground level. I do not think that ground level exists, but rather that we are faced always with the possibility of an infinite regress, which can be avoided only by making conventions about which point to start from for a particular purpose. But if one *does* accept this hierarchical view, then the theory of meaning cannot serve as the foundation, for the following rather obvious reason.

Let us assume that the problem of meaning does have some philosophical relevance. (I shall go on to question *this* assumption afterwards.) Dummett constantly stresses that what is wanted is a theory of meaning, rather than an analysis of particular meanings. Now there is always more than one (actual or possible) theory about anything, and accordingly we shall have to chose between *alternative* theories of meaning. The question which arises then is: what, if any, criteria are there which will enable us to decide which of any proposed theories is the better? Now this is a problem which is certainly prior to the actual choice of a theory of meaning, because it forms the basis of our choice. It is an epistemological problem, since it concerns the appraisal of our knowledge. This is sufficient to refute Dummett's claim that '. . . the theory of meaning [is] the only part of philosophy whose results do not depend upon those of any other part'.[2]

The dependence of theories of meaning on epistemology has been a standard pattern in philosophy. I have already given the example of Frege's theory of meaning. The same could be said of the verificationist theory of meaning, which floundered, not on results in the theory of meaning, but on negative logical and epistemological results.[3] Again, Dummett refers often to the debate between intuitionistic and classical logic, which he regards as a debate within the theory of meaning.[4] But it is very clear—from Brouwer's writings for instance[5]—that intuitionism takes its stand primarily on epistemological arguments.

There is another way in which epistemology may play a role in determining our decisions about meanings of terms. Consider again the example of classical versus intuitionistic logic. One of the main purposes of both these systems is as an organum of proof in mathematics. Now classical and intuitionistic mathematics may be viewed as rival research programmes. They conflict over many of the theorems that they generate, and their heuristics—which consist partly of their proof procedures—are quite different.[1] Therefore our appraisals of the relative merits of the two research programmes will influence our decision about which logic to accept. And the question, how can we appraise rival theories in science or mathematics, is an epistemological one.[2]

---

[1] See Lakatos [1962].

[2] Dummett [1973], p. 669.

[3] *Cf.* Popper's critique in his [1934], *e.g.* sections 4, 6, 13, 26, 79.

[4] Dummett, *ibid.*, *e.g.* p. 507.

[5] Brouwer starts from the assumption that *genuine* mathematical knowledge must be indubitable (see Brouwer [1913], p. 67). But the actual infinite must transcend human knowledge. Therefore genuine mathematics must be concerned with human constructions rather than transcendent entities, and so classical logic does not apply in mathematics. Heyting starts from the same epistemological principle: that mathematical certainty is not consistent with the assumption of transcendent entities: 'If "to exist" does not mean "to be constructed", it must have some metaphysical meaning. It cannot be the task of mathematics to investigate this meaning or to decide whether it is tenable or not' (Heyting [1956], p. 2).

This argument sheds further doubt on the philosophical relevance of meaning. Clearly, questions of meaning arise with respect to any dispute between theories. Any theory about something can also be interpreted as a theory about the meanings of the words used to describe that thing. But meaning only enters at second hand: our decisions about the meanings of the relevant expressions are dictated by our appraisals of the relative merits of the theories involved, not by a 'theory of meaning'.

I should like to mention in passing that, when Dummett gets down to saying what a theory of meaning should actually look like, he falls into an error which Frege spent a lot of time trying to expose. This is the error of giving a descriptive answer to an essentially normative question.[3] Dummett, it appears, wants to base a theory of meaning on a description of how human beings actually use, and learn to use, their native language. The following examples illustrate the recurrence of this thesis throughout the book.

> The actual fact of our linguistic practice is that the only notions of truth and falsity which we have for [an undecidable sentence] are ones which do not entitle us to regard the sentence as determinately true or false independently of our knowledge.[4]
>
> What is in question here is the correct model of meaning, of what we learn when we learn to use the sentences of our language; . . .[5]
>
> Philosophical analysis is, in a sense, an attempt to reverse the process [of learning], and, in doing so, to make explicit what had been only implicit in the learning process.[6]

But it is, I think, quite clear that a decision about meaning is, at least from a philosophical point of view, a normative rather than a descriptive matter. Having discovered (by descriptive investigation) what meaning we actually learn to ascribe to an expression, a further, philosophically more important question arises: is that assignment of meaning justified? Surely, for instance, we cannot rest content with the knowledge of how a pragmatist comes to use the word 'true' in the way that he does. Anyone who accepts the classical (Tarskian) theory will wish to argue with him; pointing out, for instance, that pragmatism leads to the peculiar conclusion that a statement can be both true and false. If we cannot enquire into the justification of usage, philosophical debate would become impossible. What 'being justified' is going to amount to can only be determined by a norm, and this Dummett fails to provide.[1]

---

[1] For an explanation of these—by now well known—technical terms, see Lakatos [1970].

[2] I refer the reader to the work of Lakatos [1976*a*], wherein a case is made out for the strong analogy between the methodology of mathematics and that of the physical sciences. Lakatos has also shown how the meanings of crucial concepts in mathematics have often developed as the result of clashes between rival (informal) mathematical theories. See Lakatos [1976*c*].

[3] See *above*, p. 85.

[4] Dummett [1973], p. 468.

[5] *Ibid.*, p. 508.

[6] *Ibid.*, p. 668. See also pp. 515 and 591.

7 There is a great deal in this book about which this review has said nothing. There are indeed things of value in it: many chapters contain pointers to problems in the fine structure of Frege's philosophy, and there are thought-provoking criticisms of Kripke, Quine, and other contemporary philosophers. There are also passages which do not conform very well to the interpretation that I have put on the book. But I think that to exploit them in Dummett's favour may only convict him on the further charge of inconsistency. For instance, while the last chapter congratulates Frege for enshrining the theory of meaning as 'first philosophy', the ***Introduction*** says something rather different; that it was Wittgenstein who overthrew the Cartesian tradition by giving pride of place to philosophical logic rather than to epistemology, and that Frege's contribution to this overthrow was to show that logic is independent of any prior philosophical considerations.[2] Later he says that logic and the theory of meaning are not the same, and that the term 'philosophical logic' is a wholly misleading one.[3] At one point he says that '. . . the general contention that Frege wanted to extrude everything epistemological from logic or from the theory of meaning is quite misconceived: he wanted to extrude everything psychological; but that was for him quite a different matter'.[4] The first part of this sentence is inconsistent with the claim (mentioned directly above) that Frege held logic to be independent of any other philosophical doctrine. Having said that a theory of meaning should describe what the learning of a language consists of,[5] Dummett later says that 'a study of [the learning process] cannot be required for an analysis of sense'.[6] All this does not make the task of extracting an unambiguous position from the book any easier.

8 In conclusion I should like to make it quite clear that I have not criticised Dummett's claim that Frege was a great philosopher, but only his arguments in favour of this claim. There are other reasons—and I have hinted at them already—for ascribing greatness to Frege. One of them was that he began a new research programme in epistemology, concentrating on knowledge as formulated in objective propositions, rather than as a special form of belief, or as an 'activity'.[7] Frege's contribution to the programme was to characterise knowledge in this way, and to argue for this characterisation. The problem that remained was to provide ways of appraising knowledge, when isolated in this way from the context of belief. This has been the central problem of modern philosophy

---

[1] It is ironic that Dummett is here violating his own expressed intention to avoid the uncritical attitude towards language that ordinary language philosophers are guilty of (see *above*, p. 79).

[2] *Ibid.*, p. xv. [3] *Ibid.*, p. 670. [4] *Ibid.*, p. 240.

[5] *Ibid.*, p. 508, quoted *above*, p. 89. [6] *Ibid.*, p. 677.

[7] Outright subjectivist theories of knowledge, such as the psychologism that Frege criticised, are more or less dead these days. But they have been replaced by theories which characterise knowledge as a certain kind of human activity, or disposition to act, and these theories have the same relativistic consequences as the earlier ones. The epistemologies of both Wittgenstein and Quine are of this nature. For the relativism inherent in Wittgenstein's theory of 'language games', see Lakatos [1976*b*].

of science,[1] and the greatest contributors to the development of methodology, Carnap and Popper, were deeply influenced by Frege.[2] In fact their work, and the work of others in this field, is the natural continuation of his.

GREGORY CURRIE
London School of Economics

REFERENCES

Brouwer, L. E. J. [1913]: 'Intuitionism and Formalism', *Bulletin of the American Mathematical Society*, **20**, pp. 81–96. Reprinted in P. Benacerraf and H. Putnam (*eds.*): *Philosophy of Mathematics: Selected Readings*, pp. 66–77. Oxford: Basil Blackwell.

Carnap, R. [1950]: *Logical Foundations of Probability*. London: Routledge and Kegan Paul.

Carnap, R. [1963]: 'The Development of My Thinking', in P. A. Schilpp (*ed.*): *The Philosophy of Rudolf Carnap*, pp. 3–43. La Salle: Open Court.

Dummett, M. [1973]: *Frege. The Philosophy of Language*. London: Duckworth.

Frege, G. [1879]: *Begriffsschrift*. English translation in J. van Heijenoort (*ed.*): *From Frege to Gödel*. Cambridge, Massachusetts: Harvard University Press.

Frege, G. [1884]: *Die Grundlagen der Mathematik*. German text with English translation by J. L. Austin. Oxford: Basil Blackwell. Revised edition 1953.

Frege, G. [1892]: 'Über Sinn und Bedeutung', *Zeitschrift für Philosophie und Philosophische Critique*, **100**, pp. 25–50. Reprinted in I. Angelelli (*ed.*): *Kleine Schriften*, pp. 143–62. Darmstadt: Wissenschaftliche Buchgesellschaft, 1967.

Frege, G. [1893]: *Grundgesetze der Arithmetik*. Hildesheim: George Olm, 1962.

Frege, G. [1906]: 'Über die Grundlagen der Geometrie, III', *Jahresberichte der Deutschen Mathematischer-Vereinigung*, **15**, pp. 423–30. Reprinted in *Kleine Schriften*, pp. 317–23.

Frege, G. [1914]: 'Logik in der Mathematik', in H. Hermes, F. Kambartel and F. Kaulbach (*eds.*): *Nachgelassene Schriften*, pp. 219–70. Hamburg: Felix Meiner Verlag, 1970.

Frege, G. [1918*a*]: 'Der Gedanke: Eine Logische Untersuchung', *Beiträge zur Philosophie des Deutsches Idealismus*, **1**, pp. 58–77. Reprinted in *Kleine Schriften*, pp. 342–62.

Frege, G. [1918*b*]: 'Die Verneinung: Eine Logische Untersuchung', *Beiträge zur Philosophie des Deutschen Idealismus*, **1**, pp. 143–57. Reprinted in *Kleine Schriften*, pp. 362–78.

Geach, P. T. [1961]: 'Frege', in G. E. M. Anscombe and P. T. Geach: *Three Philosophers*. Oxford: Basil Blackwell.

Heyting, A. [1956]: *Intuitionism: an Introduction*. Amsterdam: North Holland.

Lakatos, I. [1962]: 'Infinite Regress and the Foundations of Mathematics', *Aristotelian Society Supplementary Volumes*, **36**, pp. 155–84.

Lakatos, I. [1970]: 'Falsification and the Methodology of Scientific Research Programmes', in I. Lakatos and A. Musgrave (*eds.*): *Criticism and the Growth of Knowledge*, pp. 91–195. Cambridge University Press.

Lakatos, I. [1976*a*]: 'A Renaissance of Empiricism in the Recent Philosophy of Mathematics?', *British Journal for the Philosophy of Science*, **27**. *Forthcoming*.

---

[1] I should say 'non-Wittgensteinian philosophy of science'. Wittgensteinian philosophy of science has continued to concentrate on the old subjectivist problem of 'understanding' (see *e.g.*, Toulmin [1972], and Watson [1967]).

[2] Carnap attended Frege's lectures at Jena (see Carnap [1963]) and subsequently wrote against psychologism (Carnap [1950], p. 37–51). Popper has recently developed some of Frege's views on epistemology, though I do not think that the extent of his debt to Frege has been sufficiently recognised. The term 'Third World' (*drittes Reich*) is due to him, and he even suggested that the third world (of propositions) interacts with the first (physical) world *via* the second world of psychology. (See his [1918*a*], pp. 353–4, and 361–2. Compare Popper [1972], ch. 3 and 4.)

LAKATOS, I. [1976*b*]: 'Understanding Toulmin', *Minerva, forthcoming.*
LAKATOS, I. [1976*c*]: *Proofs and Refutations: The Logic of Mathematical Discovery.* Cambridge University Press. *Forthcoming.*
POPPER, K. R. [1934]: *Logik der Forschung.* Vienna: Springer. Translated into English as *The Logic of Scientific Discovery.* London: Hutchinson.
POPPER, K. R. [1972]: *Objective Knowledge.* Oxford University Press.
STOOTHOFF, R. H. [1963]: 'Note on a Doctrine of Frege', *Mind*, **72**, pp. 406–8.
TOULMIN, S. [1972]: *Human Understanding*, **1**. Oxford University Press.
WATSON, W. H. [1967]: *Understanding Physics Today.* Cambridge University Press.
WITTGENSTEIN, L. [1951]: *Philosophical Investigations.* Oxford: Basil Blackwell.
WITTGENSTEIN, L. [1956]: *Remarks on the Foundations of Mathematics.* Oxford: Basil Blackwell.

Brit. J. Phil. Sci. 27 (1976), 93–96 *Printed in Great Britain*

Reviews

QUINE, W. V. [1973]: *The Roots of Reference*. La Salle: Open Court. Pp. 151. $10.00.

Quine's latest book tackles a problem raised in *Word and Object*: how do we learn the linguistic apparatus—'some', 'another', 'that', 'it' and the plural 's', for example—by means of which we speak about objects, abstract, physical, or sensory? Previously Quine had offered only the metaphorical suggestion that 'the child scrambles up an intellectual chimney, supporting himself against each side by pressure against the others' ([1960], p. 93); in *The Roots of Reference* his aim is a more literal and detailed account. This account proceeds in three stages, offering answers to three questions: how, given the evidence of our senses, do we acquire our theory of the world? how, given that evidence, do we acquire mastery of 'cognitive language', observation sentences, predication, the categorical, the truth-functions? and how, more specifically, do we acquire mastery of the linguistic apparatus of reference?

Quine regards the first of the questions, however unfamiliar, as the legitimate heir of the traditional epistemological problem of scepticism, which he sees as having challenged science by pointing to the limitations of the evidence on which it rests. Quine's approach differs from that of traditional 'foundationalist' epistemologies, however, for, in its abandonment of the search of a 'first philosophy' to provide a basis for science, it can allow, without danger of circularity, for epistemology to appeal to psychology. The psychology Quine favours is behaviourist; perception is a matter of conditioning, of habit formation, of dispositions to respond to stimuli. This is refined, however, by supplementing the notions of stimulus and response with those of perceptual similarity and pleasantness. After preliminary attention to the notions of disposition and perceptual similarity, Quine offers a hedonistic model of learning in terms of a drive to increase the similarity of the present episode to a pleasant past episode, the strength of the drive varying with the vividness of the memory trace, which in turn varies with the degree of pleasantness of the earlier episode and its degree of perceptual similarity to the present one.

To account for mastery of scientific theory, Quine argues, one needs to account for mastery of scientific language. When he turns to the question of the learning of scientific language, Quine maintains what he calls his 'continuing adherence to externals'. Observation sentences are intersubjectively agreed upon (this is given as a definition of observation sentence in Section 10 as it was in 'Grades of Theoreticity') and consequently play a basic role both in support of scientific theories, and in the learning of language. The verification theory of meaning, Quine argues, was correct insofar as it recognised that observation sentences constitute evidence for theory just because of the associations between observable events and theoretical language by means of which that language is learned; where it went wrong was in treating sentences individually, instead of in the blocks to which, jointly, observations are relevant.

Observations are the 'gateway to language', for they, being directly keyed to current stimulation, can be learned by direct conditioning; pointing speeds up the process by emphasising what is salient. 'Perceptual similarity and the pleasure principle' explain acquisition of assent: the child, recalling the conspicuous patch of red, his own 'red', and his parent's approving 'yes', heightens the resemblance of the present colour and the sound 'red' by adding 'yes'. This lesson is generalised when the child appreciates that assent to a sentence entails the same rewards or penalties as repetition of the sentence—and thus, now, rests on appreciation of a language-dependent similarity. Quine then proceeds to offer an account of the mastery of the distinction between mass terms and terms for bodies, the latter being once again a matter of language-dependent learning, but this time by appreciation of similarity of similarities; then of the mastery of attributive composition, thence predication, the categorical, and eventually the eternal sentence and the truth-functions.

Beyond these 'primitive steps', in the acquisition of theoretical language, Quine turns his attention to the apparatus of reference. Mastery of quantification is explained as derived from the grasp of two more primitive idioms, the relative clause and the categorical. The relative clause, Quine suggests, is mastered by recognition of the interchangeability of 'Fa' with 'a is a thing x, such that Fx', recognition explained by the inductive discovery that speakers will assent to the former when and only when they will assent to the latter. Then, appreciating that general terms and relative clauses alike take predicate position, the learner pursues the analogy by allowing the relative clause, like the general term, into categoricals, 'Every α is a β'. The universally quantified '(x) (if Fx then Gx)' is the universal categorical 'Every α is a β' with the relative clauses 'thing x such that Fx' and 'thing x such that Gx' for 'α' and 'β'. At its inception, then, the idiom of quantification is substitutional. But, Quine argues, the idiom is then extended, analogically, until it can no longer be substitutional, but must be objectual; the categorical construction is learned by way of its instances, but then outruns them through employment in cases where the instances (apples, rabbits, grains of sand) are nameless. With this 'irreducible leap' into objectual quantification, reference to objects is achieved.

Some, I fear, will object that Quine's enterprise is not properly speaking philosophical at all, but is, rather, an exercise in speculative psychology. Those, like myself, who sympathise with Quine's *rapprochement* between philosophy and science will not regard the frequently psychological character of the book as any kind of drawback. Others, no doubt, will object to Quine's rather strictly empiricist approach to language learning; for it is currently fashionable to suppose that the question of how language is learnt should be approached indirectly, by way of the question of the structure of the language acquired. It has, indeed, even been maintained that much linguistic apparatus cannot have been learned at all, but must be innate. But the fashionable scepticism about the capacity of behaviourist psychology to account for language acquisition seems to me to add to, rather than detract from, the interest of Quine's attempt to see how much *can* be explained by induction and analogy.

Granted the legitimacy and the interest of the approach, however, one might have hoped to have been offered rather fuller suggestions about the kinds of experimental findings that would constitute tests of Quine's psychological speculations; for in the absence of such suggestions one is apt to feel uncomfortable when Quine speaks of language as acquired by 'successions of leaps', some of them 'irreducible', if 'plausibly short'. But no sooner has one felt the need for some indication how Quine thinks his speculations might be tested, than one finds him admitting that what he is offering is not *exactly*, after all, a theory about how English is in fact learned. One is alerted by Quine's comment that his account, though presented throughout as applying to the acquisition of language by the individual learner, could equally be regarded as applying to the historical development of the language itself. Since it is by no means obvious that these must, or do, proceed in parallel, this raises some questions about the status of Quine's speculations. In part, at least, it turns out, Quine calls his account 'imaginary', and a 'caricature', though 'plausible in its essentials' because what he offers an account of is not, strictly speaking, the acquisition of English, but the acquisition of a regimented language, predicate calculus, which 'bypasses the sinuosities' of natural languages. This much schematisation might have been unexceptionable—though Quine admits at one point that it has 'carried us far from the human situation'—were it not that in other places Quine dismisses certain aspects of regimented language, such as the possibility of specifying each grain of sand by spatio-temporal co-ordinates, or defining numbers set theoretically, as 'wildly irrelevant to genetic considerations' and 'absurdly remote from the typical language learner'.

And while, on the one hand, remarks like these make it less clear than one might have desired what Quine takes to be the constraints on his inquiry, others leave one in some doubt about what Quine takes to be the consequences of his speculations. Although Quine expresses the hope that an account of its psychogenesis will 'enhance our understanding of reference itself' and although he says that he is interested in the theory of language learning for the light that it may throw on the nature of scientific theory, he also explicitly denies (p. 91) that there need be, or always is, agreement between the most plausible account of the structure of a language and the most plausible account of the steps by which that language is acquired; and it seems to follow from this that one could not rely on a genetic account for any certain conclusions about the structure of the linguistic apparatus.

Despite these difficulties, *The Roots of Reference* is a very rich and rewarding book. Besides the main lines of argument already sketched, one finds valuable discussions of, for example, the feasibility of substitutional versus objectual qualification, of the thesis of the relativity of identity, of the analysis of relative pronouns, of Russell's paradox, of nominalism. Themes familiar from Quine's earlier works—including the relevance of psychology to epistemology urged in 'Epistemology Naturalised', and the constraints on ontological economy investigated in 'Steps Towards a Constructive Nominalism'—are developed and amplified (Sections 1 and 23 respectively). Sometimes the developments are surprising; in Section 21, for instance, Quine offers an account of analyticity—not to be sure, an account in terms of meaning, a notion of which Quine remains sceptical, but in genetic terms: a sentence is analytic if everybody learns that it is true by learning its words. The appeal to the linguistic community and to considerations of genesis are characteristic; so, too, is the conclusion that analyticity is a matter of degree, in that some sentences are learnt to be true, when they are learnt, by everybody, some by most speakers, some by few or none. Quine speculates, furthermore, that perhaps some of the laws of logic are, and others are not, analytic in this sense; this, he suggests, allows for disagreement over the law of excluded middle on the part of the Intuitionists. Here Quine seems rather more tolerant of disagreement in logic than he was in chapter 6 of *Philosophy of Logic*, where the deviant logician was in danger of finding his ostensible disagreement translated away ('When [the deviant logician] tries to deny the doctrine he only changes the subject' [1970], p. 81)). How feasible Quine's novel concept of analyticity turns out to be, however, will depend on how adequately the condition that a speaker 'learns that the sentence is true *by* learning its words' can be spelled out; 'learns that the sentence is true *when* he learns its words' would let in too much, and 'learns that the sentence is true *in virtue of* learning its words' seems dangerously close to the old accounts of analyticity as truth in virtue of meaning.

The Roots of Reference is written with the style and verve one has come to expect from Quine. The reader may perhaps suffer some twinges of intellectual indigestion; he is in no danger, however, of the worse pangs of intellectual malnutrition.

SUSAN HAACK
University of Warwick

REFERENCES

QUINE, W. V. and GOODMAN, N. [1947]: 'Steps Towards a Constructive Nominalism', *Journal of Symbolic Logic*, **12**, pp. 105–122.

QUINE, W. V. [1960]: *Word and Object*, Wiley, 1960.

QUINE, W. V. [1969]: 'Epistemology Naturalised' in *Ontological Relativity*, Columbia U.P., 1969.

QUINE, W. V. [1970]: *Philosophy of Logic*, Prentice-Hall, 1970.

QUINE, W. V. [1970]: 'Grades of Theoreticity' in Foster and Swanson (*eds.*): *Experience and Theory*, Duckworth, 1970.

Brit. J. Phil. Sci. 27 (1976), 97–99 *Printed in Great Britain* 97

Recent Publications

BOOKS RECEIVED

Addis, L. [1975]: *The Logic of Society*. Minneapolis: University of Minnesota Press. $12.50. Pp. 226.

Agassi, J. [1975]: *Science in Flux*. Dordrecht: D. Reidel. Hardback D.fl. 120, Paperback D.fl. 65. Pp. 553.

Apostel, L. [1974]: *Matière et Forme I—Introduction à une Epistemologie Rèaliste*. Blandijnberg: Centre National de Recherches de Logique. Pp. 454.

Apostel, L. [1974]: *Matière et Forme II—Introduction à une Epistemologie Rèaliste*. Blandijnberg: Centre National de Recherches de Logique. Pp. 452.

Barnes, J., Schofield, M. and Sorabji, R. (*eds.*) [1975]: *Articles on Aristotle: 1 Science*. London: Duckworth. £3.50. Pp. 224.

Batens, D. [1975]: *Studies in the Logic of Induction and in the Logic of Explanation*. Brugge: De Tempel. 1450 Belgian Francs. Pp. 309.

Bonelli, M. and Shea, W. [1975]: *Reason, Experiment and Mysticism in the Scientific Revolution*. London: Macmillan. £13.00. Pp. 320.

Blackburn, S. (*ed.*) [1975]: *Meaning Reference and Necessity—New Studies in Semantics*. London: Cambridge University Press. £4.95. Pp. 210.

Braun, H. J. [1974]: *Technologische Beziehungen zwischen Deutchland und England*. Dusseldorf: Pädagogischer Verlag Schwann. Pp. 204.

Broad, C. D. [1975]: *Leibniz: and Introduction*. London: Cambridge University Press. £4.75. Pp. x+174.

Bruno, G. [1975]: *The Ash Wednesday Supper*. Paris/The Hague: Mouton. D.fl. 28. Pp. 174.

Burchfield, J. D. [1975]: *Lord Kelvin and the Age of the Earth*. London: Macmillan. £10. Pp. 260.

Cannavo, S. [1974]: *Nomic Inference*. The Hague: Martinus Nijhoff. D.fl. 54. Pp. x+331.

Cooper, D. E. [1975]: *Knowledge of Language*. London: Prism Press. £1.95. Pp. 196.

Cornman, J. W. [1975]: *Perception, Common Sense, and Science*. London: Yale University Press. £11.00. Pp. vii+420.

Crosland, M. (*ed.*) [1975]: *The Emergence of Science in Western Europe*. London: Macmillan. £11.00. Pp. 201.

Elsasser, W. [1975]: *The Chief Abstractions of Biology*. Amsterdam: North-Holland. $18.75. Pp. 261.

Fay, B. [1975]: *Social Theory and Political Practice*. London: George Allen and Unwin. Hardback £3.25. Paperback £1.45. Pp. 123.

Feyerabend, P. [1975]: *Against Method*. London: New Left Books. £5.75. Pp. 339.

Filmer, P., Phillipson, M., Sandywell, B., Silverman, D. and Roche, M. [1975]: *Problems of Reflexivity and Dialectics in Sociological Inquiry—Language Theorising Difference*. London: Routledge and Kegan Paul. Paperback £3.90. Pp. 196.

de Finetti, N. [1975]: *Theory of Probability Vol. 2*. Chichester: John Wiley and Sons Ltd. £10.50. Pp. 375.

Fortenbaugh, W. W. [1975]: *Aristotle on Emotion*. London: Duckworth. £3.75. Pp. 99.

Fraser, J. T. [1975]: *Of Time, Passion and Knowledge*. New York: George Braziller. $17.95. Pre-pub., $20.00 thereafter. Pp. ix+529.

Gadamer, H. and Vogler, P. [1975]: *Philosophische Anthropologie* Band 6 and 7. Stuttgart: Georg Thieme. DM. 16.80. Band 6—pp. vii+456, Band 7—pp. vii+415.

Garan, D. G. [1975]: *The Key to the Sciences of Man*. New York: Philosophical Library. $10.00. Pp. ix+561.

Guthrie, W. K. C. [1975]: *A History of Greek Philosophy. Vol. 4. Plato—the Man and hi Dialogues—Earlier Period*. London: Cambridge University Press. £12.00. Pp. xii+603.

Hacking, I. [1975]: *The Emergence of Probability*. London: Cambridge University Press. £5.50. Pp. 209.

Harré, R. and Madden, E. [1975]: *Causal Powers*. Oxford: Basil Blackwell. £4.75. Pp. 191.

HARTKOPF, W. [1975]: *Die Dialektik in Schellings Transzendental un Identitätsphilosophie—Studien zur Entwicklung der Modernen Dialektik.* Meisenheim am Glan: Verlag Anton Hain. DM 49. Pp. 203.

HEGENBERG, I. [1975]: *Significado e Conhecimento.* São Paulo: Editora Pedagogica e Universitaria. Pp. 185.

HEGENBERG, L. [1975]: *Logica, Simbolizacao e Deducao.* São Paulo: Editora Pedagogica e Universitaria. Pp. 219.

HEWSON, A. M. [1975]: *Giles of Rome and the Medieval Theory of Conception.* London: The Athlone Press. £10.00. Pp. viii+268.

HINCKFUSS, I. [1975]: *The Existence of Space and Time.* Oxford: Clarendon Press. £3.50. Pp. 153.

HIRST, P. Q. [1975]: *Durkheim, Bernard and Epistemology.* London: Routledge and Kegan Paul. £4.95. Pp. 212.

ICERMAN, L. and PENNER, S. [1974]: *Energy,* Vol. 1. Reading, Massachusetts: Addison Wesley. Hardback $14.50. Paperback $8.50. Pp. vi+373.

JENC, F. [1973]: *Notes in Mathematical Physics: Atomicity and Maximality in Axiomatic Quantum Theory.* Marburg: Abteilung Mathematische Physik des Fachbereichs Physik der Universität Marburg. Pp. 14.

JUSSILA, E. [1975]: *Functional Logic: Sentences.* Finland: Kangasalan Kirjapaino. Pp. 3+52.

KEAT, R. and URRY, J. [1975]: *Social Theory as Science.* London: Routledge and Kegan Paul. Hardback £6.95. Paperback £3.25. Pp. 278.

KÖRNER, S. (*ed*): [1975]: *Explanation.* Oxford: Basil Blackwell. £5.50. Pp. 219.

KOUMAKIS, G. [1975]: *Karl Jaspers—Man and Technology.* Athens: J. Chiotellis. Pp. 307.

KREBS, H. A. and SHELLEY, J. H. (*eds.*) [1975]: *The Creative Process in Science and Medicine.* Amsterdam: Excerpta Medica. D.fl. 52.00. Pp. v+138.

LENARTOWICZ, P. [1975]: *Phenotype–Genotype Dichotomy: An Essay in Theoretical Biology.* Rome: Typis Pontificiae Universitatis Gregorianae. Pp. 235.

LUDWIG, G. [1974]: *Notes in Mathematical Physics: Measurement as a Process of Interaction Between Macroscopic Systems.* Marburg: Abteilung Mathematische Physik des Fachbereichs Physik der Universität Marburg. Pp. 27.

OLSON, R. [1975]: *Scottish Philosophy and British Physics 1750–1880.* New Jersey: Princeton University Press. £9.20. Pp. 349.

OUTHWAITE, W. [1975]: *Understanding Social Life.* London: George Allen and Unwin. Hardback £3.85. Paperback £1.85. Pp. 127.

POMPA, L. [1975]: *Vico: A Study of the 'New Science'.* London: Cambridge University Press. £4.90. Pp. ix+194.

RANDALL, J. [1975]: *Parapsychology and the Nature of Life.* London: Souvenir Press. £4.00. Pp. 256.

REMMLING, G. W. [1975]: *The Sociology of Karl Mannheim.* London: Routledge and Kegan Paul. £4.95. Pp. viii+255.

ROBINSON, H. J. [1975]: *Renascent Rationalism.* Toronto: Macmillan of Canada. $14.95. Pp. 290.

SALMON, W. C. [1975]: *Space Time and Motion: A Philosophical Introduction.* Encino, California and Belmont California: Dickenson. Pp. ix+147.

SEREBRIAKOFF, V. [1975]: *Brain.* London: Davis-Poynter. £6.00. Pp. 185.

SLOTE, M. [1975]: *Metaphysics and Essence.* Oxford: Basil Blackwell. £4.25. Pp. 160.

SODDY, F. [1975]: *Radioactivity and Atomic Theory.* London: Taylor and Francis Ltd. £12.00. Pp. 517.

SPISANI, F. [1975]: *Philosophical Foundations of Autogenetic Logic.* Bologna: Centro Superiore di Logica e Scienza Comparate. Pp. 115.

STEINER, M. [1975]: *Mathematical Knowledge.* Ithaca and London: Cornell University Press. £5.20. Pp. 167.

STRAWSON, P. F. [1974]: *Subject and Predicate in Logic and Grammar.* London: Methuen. Hardback £3.50. Paperback £1.50. Pp. 144.

STUDDERT-KENNEDY, H. [1975]: *Evidence and Explanation in Social Science: An Interdisciplinary Approach.* London: Routledge and Kegan Paul. £6.50. Pp. 236.

STUEWER, R. [1975]: *The Compton Effect.* New York: Science History Publications. Pp. 367.

TAYLOR, J. [1975]: *Superminds.* London: Macmillan. £3.95. Pp. 183.

THOM, R. [1975]: *Structural Stability and Morphogenesis*. Massachusetts: Addison-Wesley/W. A. Benjamin. Hardback $22.50. Paperback $13.50. Pp. 348.

TODD, W. B. (*ed.*) [1975]: *Hume and the Enlightment*. Edinburgh University Press. £5.50. Pp. 215.

VALLANCE, E. (*ed.*) [1975]: *The State, Society and Self-Destruction*. London: George Allen and Unwin. £4.20. Pp. 150.

THE CENTER FOR THE PHILOSOPHY OF SCIENCE, AND THE DEPARTMENT OF HISTORY AND PHILOSOPHY OF SCIENCE AT THE UNIVERSITY OF PITTSBURGH

The Difference Between People and Mindless Machines
George Schlesinger, Department of Philosophy, University of North Carolina
Wednesday, 1 October 1975.

A Third Dogma of Empiricism
Wesley C. Salmon, Department of Philosophy, University of Arizona
Wednesday, 15 October 1975.

Is Undergoing Psychoanalysis Essential for the Appraisal of Psychoanalytic Theory?
Irwin Savodnik, Department of Psychiatry, University of Pittsburgh
Tuesday, 18 November 1975.

Astronomy and Astrology: The Heavenly Twins
Engelbert L. Schücking, Department of Physics, New York University
Thursday, 4 December 1975.

The Sources of Conventionalism in 19th Century Geometry
J. Alberto Coffa, Department of History and Philosophy of Science, Indiana University, Bloomington
Thursday, 22 January 1976.

Scientific Discovery: Random Search or Reasoned Search?
Herbert A. Simon, Professor of Computer Science and Psychology, Carnegie-Mellon University
Wednesday, 11 February 1976.

The Truth of Science: Beyond Realism and Instrumentalism
Stephen Toulmin, Professor of Social Thought and Philosophy, University of Chicago.
Wednesday, 10 March 1976.

The Role of Theories of Scientific Method in the Scientific Revolution
Noretta Koertge, Department of History and Philosophy of Science, Indiana University, Bloomington
Wednesday, 7 April 1976.

All the meetings at the Lecture Hall, Graduate School of Public Health, 8.15 p.m. Inquiries can be made by writing to the Director of the Center, Professor Adolf Grünbaum, 2509 Cathedral of Learning, University of Pittsburgh, Pittsburgh, Pa. 15260, U.S.A.

BOSTON COLLOQUIUM FOR THE PHILOSOPHY OF SCIENCE

SIXTEENTH ANNUAL PROGRAMME, 1975–76

Is There a Logic of Empirical Sciences?
Maria Luisa Dalla Chiara, Philosophy, University of Florence
Commentators: Joseph Agassi, Philosophy, Boston University
John Stachel, Physics, Boston University
Tuesday, 9 September, 1975

The Misuse of Models in Physics
HEINZ R. POST, History, Chelsea College (London)
Commentator: EARLE L. LOMON, Physics, M.I.T.
Tuesday, 23 September, 1975

Marxian Paradigms Versus Microeconomic Structures
WERNER LEINFELLNER, Philosophy, University of Nebraska
Commentator: HERBERT GINTIS, Economics, University of Massachusetts, Amherst
Tuesday, 14 October, 1975

Competence Versus Performance in the Growth of Scientific Knowledge
SANDRA G. HARDING, Philosophy of Social Science, SUNY, Albany
Commentators: NED BLOCK, Philosophy, M.I.T.
BARBARA HALL PARTEE, Philosophy, University of Massachusetts, Amherst
Tuesday, 4 November, 1975

Effectiveness—Does the Orthodox Church View Need Re-Examination?
ROHIT PARIKH, Mathematics, Boston University
Commentator: GEORGE S. BOOLOS, Philosophy, M.I.T.
Tuesday, 18 November, 1975

Natural Selection and Human Behaviour
RICHARD C. LEWONTIN, Zoology, Harvard University
ROBERT TRIVERS, Biology, Harvard University
Commentator: NORMAN DANIELS, Philosophy, Tufts University
Tuesday, 2 December, 1975

On the Relevance of Biology to Understanding Human Social Behaviour
MICHAEL A. SIMON, Philosophy, University of Connecticut
Commentator: E. O. WILSON, Biology, Harvard University
Tuesday, 27 January, 1976

Social Science Methodology as the Ideology of Authority in Organisation. The Moral Universe of the Corporate Executive
ALASDAIR MACINTYRE, Philosophy and Political Science, Boston University
Commentator: HUGO BEDAU, Philosophy, Tufts University
Tuesday, 10 February, 1976

The Status of the Rationality Principle in the Social Sciences
SPIRO J. LATSIS, Economics and Philosophy of Science, Boston University
Commentator: NORETTA KOERTGE, History and Philosophy of Science, Indiana University
Tuesday, 24 February, 1976

The Apparition of Clonal Man
LEON EISENBERG, Psychiatry, Harvard Medical School
Commentator: DAVID SOLOMON, Philosophy, University of Notre Dame and Boston University
Tuesday, 16 March, 1976

Formal Languages and Natural Languages: A Tenable Dualism
J. M. MORAVCSIK, Philosophy, Stanford University
Commentators: JERRY FODOR, Philosophy, M.I.T.
GILBERT HARMAN, Philosophy, Princeton University
Tuesday, 23 March 1976

Tracking Down the Misplaced Concreton in the Neurosciences
ARMAND SIEGEL, Physics and Psychiatry, Boston University
Commentator: JEFFREY ROSEN, Psychology, Boston University
Tuesday, 6 April, 1976

Knowledge and Verification in the Tractatus
BRIAN F. MCGUINNESS, Philosophy, The Queen's College, Oxford
Commentator: JOHN N. FINDLAY, Philosophy, Boston University
Tuesday, 13 April, 1976

Causation in Medicine: The Disease Model
CAROLINE WHITBECK, Philosophy, SUNY, Albany
Commentator: BARBARA ROSENKRANTZ, History of Science, Harvard University
Tuesday, 27 April, 1976

Theories of Plausibility
RISTO HILPINEN, Philosophy, University of Turku
Commentators: ISAAC LEVI, Philosophy, Columbia University
SYLVAIN BROMBERGER, Philosophy, M.I.T.
Tuesday, 4 May, 1976

Regular meetings (except 23 March) to be held at 7.45 p.m. in Room 314 of the Sherman Union, 755 Commonwealth Avenue, at Boston University Bridge. For further information, consult Professor R. S. Cohen, Department of Physics, Boston University.

PHILOSOPHICAL MONOGRAPHS

Philosophical Monographs announces that it will publish monograph-length studies in philosophy beginning Spring 1976. *PM* is an independent, non-profit venture, assisted by the University City Science Center, an established non-profit consortium in Philadelphia, and devoted exclusively to servicing the publishing needs of the profession. Manuscripts are invited, of not more than 200 typescript pages, on any topic of general interest to the profession at large. These should be accompanied by adequate return postage and an explanatory brief. All inquiries should be addressed to the Editor, Professor Joseph Margolis, Department of Philosophy, Temple University, Philadelphia, Pennsylvania 19122. The Editorial Advisory Board includes: Kurt Baier (Pittsburgh), Steven Barker (Johns Hopkins), Panayot Butchvarov (Iowa), Norman Care (Oberlin), V. C. Chappell (Massachusetts, Amherst), Jerry Fodor (MIT), Donald Gustafson (Cincinnati), Michael Lockwood (Oxford), Ernin McMullin (Notre Dame), Harold Morick (SUNY Albany), Terence Penelhum (Calgary), Jerome Shaffer (Connecticut), H. S. Thayer (CUNY), Zak Van Straaten (Witwatersrand), Richard Wasserstrom (California, Los Angeles).

AUSTRALASIAN ASSOCIATION FOR LOGIC

At a conference of logicians that was sponsored by the Research School of Social Sciences, The Australian National University, and which was held in Canberra, Australia, from 22 to 24 August 1975, it was decided to formally reconstitute the *Australasian Association for Logic* (A.A.L.), which had not met for several years. The meeting elected a new executive which will serve until the next general meeting of the A.A.L. in August 1976. Mr F. R. Routley of the Department of Philosophy, Research School of Social Sciences, The Australian National University, was elected President, and Mr M. A. McRobbie of the Department of Philosophy, Research School of Social Sciences, The Australian National University, was elected Secretary.

The object of the A.A.L. is to further studies in logic, in particular in formal logic, history of logic, foundations of mathematics, philosophy of mathematics and philosophy of logic, and the application of formal logic to problems in philosophy, in the social sciences and natural sciences, and in technology. In order to advance these aims, the A.A.L. will be holding an annual conference. The 1976 annual conference of the A.A.L. will be held at The Institute of Advanced Studies, The Australian National University, Canberra, Australia, from Thursday 12 August 1976, until Saturday 14 August 1976.

Membership of the A.A.L. is open to any person who is interested in the study of logic, and the furtherance of its study. Any person who wishes to join the A.A.L. and thereby receive notices of the A.A.L.'s activities and conferences should contact the Secretary at the following address: Mr M. A. McRobbie, Secretary, Australasian Association for Logic, Department of Philosophy, Research School of Social Science, The Australian National University, Canberra, A.C.T. 2600, Australia.

Logicians and other interested persons from outside Australia or New Zealand who will be in Australia of New Zealand at the same time as the A.A.L.'s annual conference is being held and who wish to attend or participate in the conference should contact the Secretary at the above address.

Brit J. Phil. Sci. 27 (1976), 105–136 *Printed in Great Britain*

Is the Method of Bold Conjectures and Attempted Refutations *Justifiably* the Method of Science?*

by ADOLF GRÜNBAUM

1 *Introduction*
2 *Deductivistic Appraisal of Popper's Rationale for Claiming that Severe Tests Have a Greater Potential for Weeding out False Theories Than Non-Severe Tests.*
3 *Do Severe Tests Have Greater Promise Than Non-Severe Tests for More Verisimilitudinous Bolder Theories?*
  (*a*) *Preliminary Considerations.*
  (*b*) *Propositions P and* P° vis-à-vis *Qualitative Verisimilitude*
    (*i*) *Formal Considerations*
    (*ii*) *Can the Conjectures* $B_P \subseteq A_P$ *and* $E_P \subseteq N_P$ *be Warranted by Popperian Corroborations or Only Inductivistically if At All?*
    (*iii*) *Conclusions.*
  (*c*) *Popper's* Quantitative *Theory of the Content and Verisimilitude of a Hypothesis, Prop.* $P\#$, *and the Merits of Severe Tests.*
    (*i*) *Comparison of Vs. with vs.*
    (*ii*) *Popper's Probabilistic Measure of Content, and the Bearing of his Quantitative Concept of Content on his* Reductio ad Absurdum *Argument Against Probabilistic Inductivism.*
    (*iii*) *Can Popper's Quantitative Theory of Content and Vs. Give a Deductivistic Vindication of his Advocacy of Severe Tests?*

1 INTRODUCTION

Of the many seminal ideas with which Karl Popper has palpably enriched twentieth-century philosophy of science, few are more exciting than his *rationale* for viewing the method of science as being conducive to the discovery of truth. He conceives that method to be the method of bold conjectures and attempted refutations ([1972], p. 53). And he offers two major reasons for conjecturing that attempted refutations or severe

Received 5 January 1976

---

* The author thanks the National Science Foundation (U.S.A.) for the support of research.

('crucial') tests are conducive to the fruition of the scientific quest for truth:

I. The *risky* predictions of a *false* theory are more likely to fail than its *non*-risky ones ([1962]), pp. 112, 36 and 390). Hence we have rational grounds for conjecturing that severe tests have a better chance of weeding out false theories than non-severe tests.

II. Presumably it is only by quite rare good fortune that our severe tests actually issue in the attainment of true theories. And indeed the hope of such success may be utopian. Yet the 'theorem of truth-content' (Popper [1966], p. 350, Theorem 1) permits the deduction of a proposition '*P*' which furnishes a sound basis for the following claim: The results of attempted refutations of ever bolder theories may be such as to warrant the conjecture that the bolder of two theories is also the one of greater *verisimilitude*. Popper states the pertinent proposition *P* and its purported role as 'the logical basis of the method of science' in these words:

> But the stronger theory, the theory with the greater content, will also be the one with the greater verisimilitude *unless its falsity content is also greater*. [Hereafter this premiss will be called 'Proposition *P*'.]
>
> This assertion forms the logical basis of the method of science—the method of bold conjectures and of attempted refutations. A theory is the bolder the greater its content. It is also the riskier: it is the more probable to start with that it will be false. We try to find its weak points, to refute it. If we fail to refute it, or if the refutations we find are at the same time also refutations of the weaker theory which was its predecessor [footnote 22 omitted], then we have reason to suspect, or to conjecture, that the stronger theory has no greater falsity content than its weaker predecessor, and, therefore, that it has the greater degree of verisimilitude ([1972], p. 53).

In the present sequel to my [1976*a*], I shall examine the two stated constituents I and II of Popper's *rationale* for deeming the method of attempted refutations or severe tests to be conducive to the discovery of truth. As will become clear below, my treatment of II is not unmindful of the very valuable prior critiques of Popper's theory of verisimilitude by Pavel Tichý ([1974]), John H. Harris ([1974]) and David Miller ([1974]). And my overall conclusion will be that Popper has not given us a deductivistically viable *rationale* for advocating severe tests as epistemically more promising than non-severe tests. It will then remain an open question whether such a deductivistic *rationale* can be given at all, although the suspicion insinuates itself that it may be unavoidable to resort to inductivistic devices, if the widespread *intuitive* valuation of severe tests is to receive *epistemological* sanction.

2 DEDUCTIVISTIC APPRAISAL OF POPPER'S RATIONALE FOR CLAIMING THAT SEVERE TESTS HAVE A GREATER POTENTIAL FOR WEEDING OUT FALSE THEORIES THAN NON-SEVERE TESTS

Popper ([1962], pp. 112, 36 and 390) contends that the *risky* predictions of a *false* theory are more likely to fail than its *non*-risky ones. This contention is the core of his *avowedly deductivistic* rationale for conjecturing that severe tests have a better chance of weeding out false theories than non-severe ones. And since he conceives of the scientific enterprise as a search for truth, the latter conjecture, in turn, serves as one of his major reasons for advocating severe or 'crucial' tests of bold theories.

But on what grounds does he contend that the *risky* predictions of a *false* theory are more likely to fail than its non-risky ones? And do these grounds pass deductivistic muster? Popper explains:

> A theory is tested not merely by applying it, or by trying it out, but by applying it to very special cases—cases for which it yields results different from those we should have expected without that theory, or in the light of other theories. In other words we try to select for our tests those crucial cases in which we should expect the theory to fail if it is not true ([1962], p. 112).

Note Popper's claim that—on the basis of background theories—we should expect the [new] theory to fail [a 'crucial' test] *if it is not true*. As against this assertion, I say that what is warranted *deductivistically* is the following:

(*a*) The older theories (*cum* initial conditions) deductively predict an outcome contrary to the result C deductively predicted by the new theory H, or the older theories are at least deductively non-committal with respect to C, thereby making the prediction of the new theory 'risky' in *that special sense*.

(*b*) If the new theory H does *not* fail the crucial test but passes it, clearly nothing follows deductively about *its* truth. What does follow from the observation statement C is the truth of the 'infinitely' weaker disjunction D of *all* and only those hypotheses which individually entail C. For C itself is one of the disjuncts in the infinite disjunction D. And thus C entails D in addition to being entailed by D.

Note that for any true disjunct in D, the latter has an infinitude of *false* disjuncts which are pairwise *incompatible* with that true disjunct in D and, of course, also incompatible with the background knowledge. Yet—contrary to the background knowledge—each of this infinitude of false theories predicts C no less than H does, and indeed H may be false as well. Hence of what avail is it to Popper, *qua deductivist*, that by predicting C, H is one of an infinitude of theories in D incompatible with those *particular* theories with which scientists had been working by way of historical accident, and that scientists happened to have thought of the particular disjunct *H* in *D*?

According to Popper's definition of the term 'severe test', the experiment *E* which yielded the riskily predicted *C* does qualify as a 'severe test' of *H*. But surely the fact that *H* makes a prediction *C* which is incompatible with the prior theories constituting the so-called 'background knowledge' *B* does *not* justify the following contention of Popper's: A *deductivist* is entitled to expect the experiment *E* to yield a result *contrary* to *C*, unless *H* is true.

Popper ([1959], p. 419) had reproached Bacon, Whewell, Mill and eliminative inductivists generally for having run afoul of the following fact: The elimination of some competing theories by *modus tollens* still always leaves us with an *infinitude* of rival theories. But ironically, in virtue of the infinitude of false rival theories deductively relevant to any finite body of severe test outcomes, there seems to be no rational basis on which a *pure deductivist* can even hope for a good chance that *severe* tests will weed out *false* theories any better than *non*-severe tests. And hence *I cannot see any purely deductivistic rationale for advocating severe tests as being conducive to the survival of true theories.* Elsewhere (Grünbaum [1976*b*], section 3, parts III and IV), I have discussed the import of this criticism for Popper's thesis that the rationale of his theory of corroborations is deductivistically superior to that of neo-Baconian *eliminative inductions*.

But here I wish to show just how the reasoning in Popper's cited defence of 'crucial' tests is *of a piece* with the *Bayesian* justification of such tests offered by the proponent of *inductive* mathematical probabilities of hypotheses which are, of course, anathema to Popper. Wesley Salmon has put that inductivist justification very cryptically as follows:

> A hypothesis risks falsification by yielding a prediction that is very improbable unless that hypothesis is true. It makes a daring prediction, for it is not likely to come out right unless we have hit upon the correct hypothesis. Confirming instances are not likely to be forthcoming by sheer chance ([1966], p. 119).

Let me now elaborate on Salmon's concise statement by illustrating with the aid of a single example just how his assertion is legitimated by Bayes' theorem, when the latter is applied to a test which not only qualifies as severe for Popper, but also issues in the *success* of one kind of risky prediction. In accord with Reichenbach's and Salmon's probability notation, I shall let '$P(X, Y)$' denote the probability *from X to Y*, i.e., the probability of *Y on* given *X*. Since the assumption of an initial condition relevant to the universal law-hypothesis *H* is taken for granted, that initial condition will be taken to be a logical constituent of the background knowledge B. Let 'C' denote the assertion of the test-phenomenon which is predicted deductively by the conjunction B and H. And assume that this phenomenon occurs. Consider the kind of C which is risky in the sense of predicting a fact that is 'novel' with respect to B, *i.e.*, a fact which is deductively independent of B rather than logically contrary to B (*cf.* Popper [1962], p. 390). Then $P(B, C)$ is presumably *non*-zero though small ($\leqslant \frac{1}{2}$). And in that case the form of Bayes' theorem pertinent to Salmon's statement is the following (Salmon [1966], p. 116):

$$\begin{aligned} P(B \,\&\, C, H) &= \frac{P(B, H)}{P(B, C)} \cdot P(B \,\&\, H, C) \\ &= \frac{P(B, H) \cdot P(B \,\&\, H, C)}{P(B, H) \cdot P(B \,\&\, H, C) + P(B, \sim H) \cdot P(B \,\&\, \sim H, C)} \end{aligned} \qquad (1)$$

Since we are given that $P(B \,\&\, H, C) = 1$, we can write both

$$P(B \,\&\, C, H) = \frac{P(B, H)}{P(B, C)} = \frac{P(B, H)}{P(B, H) + P(B, \sim H) \cdot P(B \,\&\, \sim H, C)}, \qquad (2)$$

and

$$P(B, C) = P(B, H) + P(B, \sim H) \cdot P(B \,\&\, \sim H, C). \qquad (3)$$

Let us see now how a Bayesian is able to *justify* the following claim: After the success of an initially risky prediction becomes known, the successful outcome C predicted by H can rightly be said to have been very improbable unless H is true. Thus, I am *not* having the Bayesian simply *assume* gratuitously in advance that the value of $P(B \,\&\, \sim H, C)$ is very low. Instead, I am having him assume that a bold hypothesis H daringly but successfully predicted C. And I wish to show that unlike Popper, the Bayesian can then appeal to Eq. (2) to deduce correctly the assertion—made by Popper and Salmon alike—that the value of $P(B \,\&\, \sim H, C)$ is very low.

Since we are given a *bold* hypothesis H, we know that $P(B, H)$ is quite low, though presumably non-zero despite the universality of H. (I defer until section 3 (*c*) a discussion of Popper's treatment of the *a priori* probabilities of universal hypotheses.) Hence, for specificity, assume that $P(B, H) = 1/100$, so that $P(B, \sim H) = 99/100$. The assumption of *riskiness* for the prediction C made by H under initial conditions included in B translates into a low value of $P(B, C)$. But Eq. (3) tells us that $P(B, H) \leqslant P(B, C)$, while Eq.(2) shows that $P(B, H) < P(B, C)$ except in the very atypical limiting case of $P(B \,\&\, C, H) = 1$. Since we assumed that $P(B, \sim H) \neq 0$, it follows from Eq. (3) in the limiting case of $P(B, H) = P(B, C)$ that $P(B \,\&\, \sim H, C) = 0$, which is the lowest value possible.

In order to make the typical case of $P(B, H) < P(B, C)$ illustratively specific, suppose that $P(B, C) = 1/50$, which is twice our assumed value of $P(B, H)$. It then follows from Eq. (3) that $P(B \,\&\, {\sim} H, C) = 1/99$, a rather low value as claimed by Salmon.

But note that although the latter value is quite low, Eq. (2) yields only $\frac{1}{2}$ for the corresponding value $P(B \,\&\, C, H)$ of the posterior probability of H. Thus, even when the assumed falsity of H makes the phenomenon C quite improbable, the occurrence of C need *not* make the truth of H more probable than its falsity.

These considerations would seem to show that while Popper failed to furnish a deductivistically sound rationale for his claim that $P(B \,\&\, {\sim} H, C) \ll \frac{1}{2}$, the Bayesian is able to underwrite cogently the corresponding assertion made by the inductivist Salmon, if he is granted the two prior probabilities which codify Popper's conditions of a bold H and a risky C.

3 DO SEVERE TESTS HAVE GREATER PROMISE THAN NON-SEVERE TESTS FOR MORE VERISIMILITUDINOUS BOLDER THEORIES?

(*a*) *Preliminary Considerations*

It is vital that we distinguish the several quite different versions of Proposition P which Popper invokes as a basis for espousing severely testing the bolder of two competing theories. Indeed he appeals to these distinct versions amid employing the terms 'bolder' and 'competing' in various senses as well.

I must now articulate some of these distinctions in order to set the stage for my subsequent inquiry.

I shall use the notation of my ([1976a]) as follows: For any given theory T, '$LC(T)$' denotes Popper's non-metrical *logical content* of the theory T, which is familiar as the Tarskian consequence class of T; '$Ct(T)$' will denote Popper's ([1972], pp. 52–3) *generalised* non-metrical content of any given T, which is presumably such that $Ct(A) \subset Ct(B)$ iff *either* $LC(A) \subset LC(B)$ *or* $QC(A) \subset QC(B)$, where $QC(T)$ is the non-metrical *question*-content $QC(T)$ of my [1976*a*]; the *capital* 'C' in '$Ct(T)$' serves to denote Popper's *unmeasured* generalised content $Ct(T)$, because I want to follow Popper in employing '$ct(T)$' with the lower case c to denote his content-*measure* of T. Again, following Popper ([1972], p. 52), we shall denote Einstein's and Newton's theories of gravitation respectively by 'E' and 'N'.

In the quotation from Popper ([1972], p. 53) given under II of I above, he speaks of a theory B as 'stronger' or 'bolder' than a theory A in each of the following two cases: (1) B *unilaterally* entails A, *i.e.*, $B \rightarrow A$ while ${\sim}(A \rightarrow B)$, so that $LC(A) \subset LC(B)$, and (2) $QC(A) \subset QC(B)$, so that

$Ct(A) \subset Ct(B)$, even when $LC(A) \not\subset LC(B)$ because A and B are logically incompatible like E and N. Moreover, he believes himself entitled to claim that whenever B unilaterally entails A, $ct(A) < ct(B)$, a supposition which I shall criticise as unfounded in 3 (*c*) below despite its *prima facie* plausibility. And in the case of the *logically incompatible* theories E and N, he claims ([1972], p. 53) not only that $Ct(N) \subset Ct(E)$ on the purported ground that $QC(N) \subset QC(E)$, but *also* that $ct(N) < ct(E)$. I shall comment on the pretentiousness of the latter grandiose *metrical* assertion in 3*c* (*ii*) below.

When Popper is concerned to compare the falsity contents of two theories A and B with a view to invoking Prop. P, he speaks of A and B as '*competing*' ([1972], pp. 52–3, and 59) not only when these theories are logically incompatible but also when one of them unilaterally entails the other. Since the logical incompatibility of two theories is a symmetric relation between them, they are competitors of one another in the proper literal sense of the term 'competing'. But in what sense can A and B be 'competing' theories when they are not only *compatible* but when one of them—say B—unilaterally entails the other? In the latter case, the proponent of A would presumably regard the logically stronger B as a 'competitor' of his A, if he is either reluctant to accept the stronger B or even rejects it. But the defender of B must clearly regard the weaker A as a mere sub-theory rather than as a 'competitor'. Thus, while someone who affirms A can consistently deny B, the converse is obviously *not* the case. Given this asymmetry between the *compatible* theories A and B, they can be said to be 'competing' with Popper in only a Pickwickian sense.

When I am concerned to distinguish Popper's two quite different cases of theory-competition from one another, I shall speak of a pair of theories as being 'incompatible competitors' or 'compatible competitors' respectively. But as is well-known from the literature on purported cases of the 'reduction' of one theory to another, cases of *bona fide* deductive subsumption of one theory under a successor theory—as distinct from merely *approximate* empirical vindication of the predecessor theory by its successor—are at best few and far between in the growth of science. Hence cases of 'competition' among compatible scientific theories are quite atypical exceptions, while competition among logically incompatible theories is typically the rule.

We can now formulate symbolically each of the several versions of the above verbal Prop. P which are invoked by Popper as a basis for his advocacy of severely testing the bolder of two theories.

For any two theories A and B, let '$A <_{vs.} B$' be short for 'A is *qualitatively* less verisimilitudinous than B' in the sense of Popper ([1962],

p. 233; [1972], p. 52), while 'A_T' and 'A_F' respectively denote the truth and falsity contents of A.[1] In the preceding issue of this *Journal*, I have noted (Grünbaum [1976*a*], section 2) how—in the case of compatible

---

[1] Let the class of all statements (of our language) be divided into the subclasses T (true) and F (false). Instead of using the symbol '$LC(A)$' of our Section 3(*a*) *above* to denote the Tarskian consequence class or content of a theory (statement) A, let us follow Popper ([1972], p. 330) and denote that Tarskian content by '$Cn(A)$'.

Popper ([1972], p. 49) defines the truth content 'A_T' of A as

$$A_T = Cn(A) \cap T. \tag{1}$$

In the corrected 1973 printing of his ([1972], p. 49), Popper defines 'the relative content of the statement *a*, *given* the content Y'—which he denotes by 'a, Y'—as 'the class of all statements deducible from a in the presence of Y, but not from Y alone' (see also his p. 332). This definition supplants the unfortunate earlier phrase 'or with the help of Y' of the 1972 printing of p. 49 by the phrase 'but not from Y alone'. Thus, if A and B are statements (theories), the definition of the *relative* content A,B of A given B is

$$A,B = Cn(A \,\&\, B) - Cn(B), \tag{2}$$

i.e., the difference set of the specified content classes. But even that corrected printing of p. 49 then still wrongly characterises 'the relative content a,B' as 'that part of the content of a [*sic*!] which goes beyond B', instead of speaking of 'that part of the content of the *conjunction* $a.b$ [assuming that B is axiomatisable by b] which goes beyond B'.

Popper avowedly ([1972], pp. 48–9 and 331) wished to define the *falsity content* A_F of A such that A_F would then achieve his stated desideratum of qualifying as a Tarskian deductive system. And he *argues* ([1972], pp. 49–50) that this aim of definining A_F so as to achieve closure of A_F under the operation of logical consequence is actually accomplished by defining A_F as the *relative* content A, A_T, *i.e.*, by the definition

$$A_F = A,A_T \tag{3}$$

But the exposition of Popper's ideas here is greatly handicapped by the fallacies which vitiate the specific arguments (*a*) and (*c*) ([1972], pp. 49–50) offered by him in support of his contention that the A_F defined in (3) is a Tarskian deductive system. Hence it may help to note that Popper repeatedly equates the assignment of *measure zero* to a subset with the *set-theoretical emptiness* of that subset. Thus he overlooks that the class L of all logically true statements is surely not empty set-theoretically, even though it is assigned the content *measure* zero no less than a set-theoretically empty content is assigned content-measure zero. According to his ([1972], pp. 49, 47 and 331), the non-empty class L is indeed a subset of the absolute content $Cn(X)$ of any theory X. But L is *not* a subset of the relative content X,L. Hence we have the *non*-identity

$$Cn(X) \neq X,L.$$

But Popper mistakenly asserts the corresponding identity ([1972], p. 49 middle; p. 50, lines 1 and 2). Although it is wrong to claim identity for the absolute content $Cn(X)$ with the relative content X,L, it *is* correct to claim equality for their content-*measures*, since the measure of L is taken to be zero.

Popper invokes the erroneous claim that $Cn(X) = X,L$ to buttress his false assertion (*a*) on p. 49 that his A_F qualifies as a Tarskian content. And the conflation of set-theoretical emptiness with measure zero issues in his assertion (*c*) on p. 50 that 'A_F "*contains*" no true statement in the sense that its [*sic*!] true statements are not taken as content, but as its (relative) zero content'. Contrary to what Popper says here, the sense in which A_F, as defined by him, *fails* to contain any true statements is the *literal* set-theoretical sense.

Once these fallacies are avoided, it is immediately evident from Popper's (1), (2) and (3) above that, as noted by Tichý ([1974], p. 156)

$$A_F = Cn(A) \cap F, \tag{4}$$

even though Popper (pp. 48–9) had rejected (4) as a definition of A_F.

'competitors' A and B!—Popper can indeed validly deduce the following version of Prop. P by means of his 1966 theorem on truth-content:

$$(\text{Prop. } P) \quad \{LC(A) \subset LC(B) \text{ and } B_F \subseteq A_F\} \rightarrow A <_{vs.} B.$$

But there (*ibid.*) I also explained why Popper failed in the case of *incompatible* competitors such as E and N as follows: His assertion

$$(\text{Prop. } P^\circ) \quad \{Ct(N) \subset Ct(E) \text{ and } E_F \subseteq N_F\} \rightarrow N <_{vs.} E,$$

which rests on what I there (*ibid.*) called 'Prop. P^*', is *at best unfounded*, even if (unbeknownst to us), E is actually true while N is false. Yet Popper invokes this gratuitous Prop. P° and the highly dubious premiss $QC(N) \subset QC(E)$ (see my [1976a, 3]) to claim ([1972], p. 53) that if appropriate results of severe tests would warrant the conjecture $E_F \subseteq N_F$, then we can *also* conjecture that $N <_{vs.} E$. Thus, he singled out E and N as a major illustration of the following thesis of his: If logically incompatible theories can be rank-ordered with respect to content Ct—which Popper claims to be possible for E and N—then Prop. P° can be legitimately invoked as a basis for advocating attempted refutations of the bolder of the two theories. Therefore, by appealing to both Prop. P° and Prop. P, he thinks he can justify his espousal of severe tests in the scientifically typical case of incompatible competitors no less than in the atypical situation of 'competition' between compatible theories.

Popper uses '$ct_T(A)$' and '$ct_F(A)$' respectively to denote his *measures* of the truth-content and falsity-content of a theory A whose *measured* content is $ct(A)$. Let '$A <_{Vs.} B$' be short for 'A is *quantitatively* less verisimilitudinous than B' in the sense of Popper ([1962], p. 234), where it is to be noted that here I use a *capital* 'V' in the abbreviation '*Vs*'. Then for any two theories X and Y the *quantitative* version $P\#$ of Prop. P states:

$$(\text{Prop. } P\#) \quad \{ct(X) < ct(Y) \text{ and } ct_F(Y) \leqslant ct_F(X)\} \rightarrow X <_{Vs.} Y.$$

In 3 (*c*) below, I shall maintain that Prop. $P\#$ is at best gratuitous not only when X and Y are incompatible competitors but even when they are compatible 'competitors'.

Clearly, either the qualitative or quantitative version of Popper's theory of verisimilitude or both can be fruitfully assessed in their own right, quite apart from the appraisal of anyone's invocation of Props. P, P° and $P\#$ as a basis for advocating severe tests. And just such fruitful critical assessments were made by both Tichý ([1974]) and Harris ([1974]). David Miller's sophisticated and detailed critique ([1974]) of Popper's qualitative verisimilitude comments on Popper's appeal to Prop. P only to the extent of making the following point (p. 168): In the case of *compatible* competitors (pp. 167 and 171), the truth of Prop. P is *unavailing* for vindicating the quest for falsifications. And Miller's reason is as follows: When one theory unilaterally entails another, the subset condition concerning falsity-content in the antecedent of Prop. P is met only in the case of theories which are both true, a case for which 'we shall hardly need the sophistications of verisimilitude', since the logically stronger true theory would then be clearly preferable (Miller [1974], p. 174).

It is indeed true that Prop. P, though proven for *compatible* competitors, is *unavailing* for vindicating severe tests, if the logically stronger of these 'competitors' is actually *false*: As Miller ([1974], p. 168) points out, in this case the stronger of the two theories will also be the one of greater falsity-content, in contravention of the subset condition $B_F \subseteq A_F$ in the antecedent of Prop. P. But, as we noted, compatible competitors are few and far between in actual science. Hence I am concerned to go beyond Miller by dealing with the scientifically cardinal case of *incompatible* competitors in order to make the following more important point: Since both Prop. P° and Prop. $P\#$ are at best unfounded, and since the very relevance of Popper's theory of verisimilitude to his purportedly deductivistic vindication of severe tests in the scientifically typical case is predicated on either Prop. P° or Prop. $P\#$, Popper's reliance on verisimilitude to furnish this vindication is *abortive at the very outset*. Hence in the light of actual science, the difficulty with Popper's appeal to the prospect of increased verisimilitude as a justification of his advocacy of attempted refutations is not so much that Prop. P is *unavailing* though true, but rather that Props. P° and $P\#$ are at best ill-founded. Indeed even when a *false* B unilaterally entails A—a case in which Prop. P is known to be *unavailing*, since then $A \nless_{vs.} B$ because $A_F \subset B_F$—Prop. $P\#$ will turn out to be gratuitous and hence cannot serve to vindicate severe tests any more than Prop. P, although it is *not* demonstratedly *precluded* in this case that $ct_F(B) = ct_F(A)$.

In sections (*b*) and (*c*) below, I shall develop the above themes in the context of doing each of the following: (1) Show that all of the basic defects of Popper's theory of *qualitative* verisimilitude can be demonstrated by the most elementary logical devices familiar to a beginning student of Venn diagrams, a result which makes it puzzling that this theory was not seen to be untenable well before Tichý, Harris and Miller published their much more technical results. (2) Point out how Popper must rely on induction by enumeration rather than on Popperian corroborations in his attempt to provide a warrant for conjecturing that the respective conditions $B_F \subseteq A_F$, $E_F \subseteq N_F$ and $ct_F(Y) \leqslant ct_F(X)$ in Props. P, P° and $P\#$ are met. (3) Present a critique of Popper's theory of *quantitative* content and verisimilitude as well as of Prop. $P\#$, a critique whose results are not redundant with those of others, so far as I know.

My conclusion will be that neither the qualitative nor the quantitative version of Popper's theory of verisimilitude can vindicate his espousal of severe tests as epistemically more significant than non-severe tests.

(*b*) *Propositions P and P°* vis-à-vis *Qualitative Verisimilitude*

(*i*) *Formal Considerations.*

Popper's qualitative definition of lesser verisimilitude ([1962], p. 233; [1972], p. 52) as rendered by Pavel Tichý ([1974], p. 156, Prop. 2.3) is as follows:

(*Def.*) 'A has less verisimilitude than B just in case either $A_T \subset B_T$ and $B_F \subseteq A_F$ or $A_T \subseteq B_T$ and $B_F \subset A_F$'.

Tichý proves the following proposition (his 2.4 on p. 156) for two theories A and B: 'If B is false then A does not have less verisimilitude than B'.[1] Hence if A *does* have less verisimilitude than B, then B is true. In other words, as David Miller ([1974], p. 170) also proved, B is closer to the truth than A *only* if B is true. Consequently, no false theory is closer to the truth than any other theory. Evidently, Tichý's Prop. 2.4 covers all three of the following mutually exclusive special cases: (*a*) a false theory B *entails* A, (*b*) a false B is *logically incompatible* with A, and (*c*) a false B does *not* entail either A or its *negation*, *i.e.*, A is *logically independent* of the false theory B.

Tichý's Prop. 2.4 encompasses these three cases with admirable elegance and conciseness. But it will be instructive to see below that precisely this virtue obscures the following fact: In the special cases (*a*) and (*b*)—which are the ones of primary interest for the growth of scientific knowledge—certain particular features of these cases are *sufficient* to make Tichý's Prop. 2.4 *immediately* and *elementarily* evident. And it will then be seen to be surprising that Popper wrongly denied the Prop. in case (*a*) as well as in the cardinal instance of case (*b*) constituted by E and N. Moreover, it will be useful to call attention to some equally obvious results pertaining to some of the cases of logical independence.

Recall that I use the arrow symbol "→" to denote logical entailment.

Case (*a*): *A false theory* B *entails* A

Here we have two subsidiary cases: (*i*) the interesting one of *one*-way entailment in which $B \rightarrow A$ but $\sim (A \rightarrow B)$, (*ii*) the dull case in which A and B are logically equivalent. In the interesting subsidiary Case (*ai*), it will be seen to be *evident* that Tichý's result 2.4 follows from the *mere* fact that B is false and that $\sim (A \rightarrow B)$! And it will be seen to be immaterial whether $B \rightarrow A$ or not.

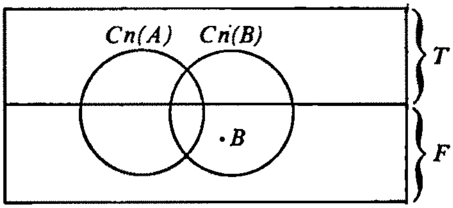
---

[1] The statement of Tichý's proof of this Prop. 2.4 contains two confusing misprints in line 5 on p. 157: 'A_F' there should read 'B_T', and 'A_T' should read 'B_F'.

(*i*) The premisses in Case (*ai*) are:

(I) $B \in F \cap Cn(B)$, *i.e.*, $B \in B_F$
(II) $B \rightarrow A$, *i.e.*, $Cn(A) \subseteq Cn(B)$
(III) $\sim (A \rightarrow B)$, *i.e.*, $B \notin Cn(A)$

If we enter the information contained in premisses (I) and (III) in a Venn diagram *without* invoking premiss (II), we obtain the following diagram:



Thus (III) and $Cn(A) = A_T \cup A_F$ yield

(III*a*) $B \notin A_F$.

And *without invoking* (*II*), we can deduce from (I) and (III*a*) that

(IV) $B_F \nsubseteq A_F$.

Recall our statement *Def.* of Tichý's rendition of Popper's definition which asserts a necessary and sufficient condition for A to have less verisimilitude than B. It is evident that our result (IV) *violates* this necessary condition, which has the form of a disjunction of two conjunctions: Our result $B_F \nsubseteq A_F$ violates both disjuncts by violating one of the two conjuncts in *each* of the two disjuncts. And note that this violation occurs, no matter whether the inclusion of A_T in B_T is proper ($A_T \subset B_T$) or improper ($A_T = B_T$) as well as when $A_T \nsubseteq B_T$.

In short, for this special Case (*ai*), Tichý's Prop. 2.4 follows *at once* from the *mere* fact that B is false and that $\sim (A \rightarrow B)$, regardless of whether B entails A or not! Recall that we write '$A <_{vs.} B$' for 'A has less *qualitative* Popperian verisimilitude than B', and note that the lower case '*v*' in '*vs.*' is used for qualitative verisimilitude, while the capital '*V*' is used in '*Vs.*' for the *quantitative* version.

This analysis enables us to comment on Popper's treatment of the following example of his. Let the actual clock time now be 9:48, and suppose that B is the assertion 'It is now between 9:45 and 9:48' while A is the assertion 'It is now between 9:40 and 9:48'. Clearly, B entails A but *not* conversely. Speaking of these two statements, Popper writes:

> If, on the other hand, we interpret 'between' so as to exclude the upper bound, then both statements become false (though they might be called 'almost true'); but they remain comparable (in the non-measure sense), and we can still assert that the first has a greater truth-likeness than the second ([1972], p. 56).

Obviously this example—*like any instance in which a false* B *entails* A *while* $\sim$ (A $\rightarrow$ B)—satisfies the simple conditions (I) and (III) from which we deduced via (IV) that $A \nless_{vs.} B$ in a most evident and elementary way. Hence it is immediately evident without the use of Tichý's much stronger Prop. 2.4 that the statement by Popper which we just quoted is incorrect. And it is somewhat surprising that the result demonstrated here for *any* Case (*ai*) was not noticed almost immediately after Popper's ([1962]) publication of his definition of $A <_{vs.} B$.

Now consider the situation in which all *three* premisses (I), (II) and (III) of our Case (*ai*) are asserted to hold.

Then from II, we have
both

(II*a*) $A_T \subseteq B_T$

and

(II*b*) $A_F \subseteq B_F$

And from (III) we have

(III*a*) $B \notin Cn(A)$, so that $B \notin A_T$ and $B \notin A_F$

Conjoining (III*a*) with (I), we have

(IV) $B \notin A_F$ and $B \in B_F$.

But from (II*b*) and (IV), we deduce

(V) $A_F \subset B_F$.

Hence B exceeds A in falsity-content as pointed out concisely by Miller ([1974], p. 171).

The result for Case (*ai*) can likewise be seen at once to cover the special case in which a false B is *logically independent* of A. To say that B is logically independent of A is to say that both $\sim (A \rightarrow B)$ and $\sim (A \rightarrow \sim B)$. Hence a false B which is logically independent of A satisfies the *sufficient* conditions of our Case (*ai*) for $A <_{vs.} B$.

Since logical independence is *not* a symmetric relation, the special case of a false B which is logically independent of A must be *distinguished* from the one in which A is logically independent of a false B; the *latter* is Case (*c*) of our tabulation.

(*ii*) A and B are logically equivalent.

In this dull case, $A_T = B_T$ and $B_F = A_F$, which violates both $A_T \subset B_T$ and $B_F \subset A_F$ in *Def.* above. Hence $A \nless_{vs.} B$ in this case.

Case (*b*)*: A false* B *is logically incompatible with* A

This, of course, is the following scientifically quite *typical* case of ***rival*** theories: A (one hopes) consistent *false* theory B and another consistent theory A (which may be either false or true) such that A and B are *logically incompatible.* If we may presume Einstein's theory to be false, then this case is instantiated by the *competing* theories E and N singled out by Popper ([1972], pp. 52–3).

My impending proof of $N \nless_{vs.} E$ can be given in a *nutshell* by three very simple steps as follows:

1. Since E is incompatible with N, we have $E \notin Cn(N)$, so that $E \notin N_F$.

2. But also $E \in E_F$, because assumedly E is false.

3. Hence $E_F \nsubseteq N_F$, which violates a necessary condition for $N <_{vs.} E$.

Let me now consider our present Case (*b*) more fully.

Clearly, the assumed falsity of B *and* its logical incompatibility with A still allow A to be either false or true, since A and B are assumed to be merely *contraries* rather than contradictories. Hence the conclusion of the proof below will pertain, like Tichý's stronger result, to a theory A which is either true or false, while B is assumed to be false. We now prove $A \nless_{vs.} B$ for this Case (*b*). We are *given* the following three Roman-numeraled conditions:

(I) B is false, so that $B \in B_F$ and $\sim B \in T$.

(II) B is logically incompatible with A, which is logically equivalent to A's logical incompatibility with B, since
$(B \rightarrow \sim A) \longleftrightarrow (A \rightarrow \sim B)$.

(III)

(*a*) B is consistent, so that $\sim B \notin Cn(B)$.

(*b*) A is consistent, so that $[\sim B \in Cn(A)] \rightarrow [B \notin Cn(A)]$.

From $A \rightarrow \sim B$ in (II), we have

(IV) $\sim B \in Cn(A)$.

Now (I) and (IV) yield $\sim B \in Cn(A) \cap T$, or

(V) $\sim B \in A_T$.

But (III*a*) entails that $\sim B$ is not a member of *any* subset of $Cn(B)$, so that

(VI) $\sim B \notin B_T$.

From (V) and (VI), we deduce

(VII) $A_T \nsubseteq B_T$.

But (VII) violates a *necessary* condition for $A <_{vs.} B$. Thus we have made immediately self-evident the deducibility of the following *weaker* form of Tichý's Prop. 2.4:

If B *is false, and* A *and* B *are logically incompatible, then* A *(whether false or true) does not have less qualitative verisimilitude than* B.

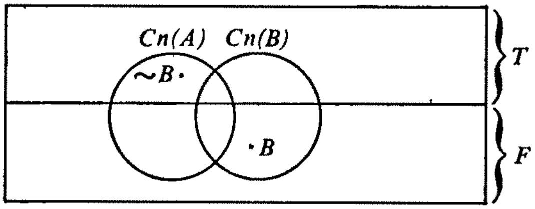
It may be noteworthy that steps (I), (IV) and (III*b*) of the proof just given yield the further conclusion that

(VIII) $B \in B_F$ and $B \notin A_F$,

so that

(IX) $B_F \nsubseteq A_F$.

Let us depict our combined results (V), (VI) and (VIII) in a Venn diagram:



Then note that the combination of (VII) and (IX) shows the following: When a false theory B is logically incompatible with a theory A, these theories violate *every* conjunct in the disjunction of conjunctions which constitutes the *definiens* in Tichý's rendition of Popper's definition of $A <_{vs.} B$. This result points up the *extent* of the revision which Popper's qualitative criterion would require if it is to handle the scientifically cardinal case of incompatible false theories.

If E and N are each false, as Popper ([1972], p. 53, n. 22) explicitly presumes, it follows from my above weaker form of Tichý's Prop. 2.4 that $N \nless_{vs.} E$.[1] It is most puzzling that Popper argues for the opposite conjecture $N <_{vs.} E$ on that page even though only a few pages later (p. 56), he states without proof *à propos* of two false statements that two incompatible statements are qualitatively *non*-comparable with respect to verisimilitude.[1] Furthermore, as Tichý ([1976] section 4) has remarked, it is unclear why Popper's avowal of the latter conclusion of non-comparability did not give him pause in regard to the adequacy of his explication, instead of prompting him to say in effect (p. 56): So much the worse for the *intuitive* impression that one of two incompatible false statements *can* be closer to the truth than another.

---

[1] In Popper's n. 22 of his ([1972], p. 53), which is appended to the long opening quotation from him given in 1(II) *above*, he tells us that as of the present, tests of the 'eclipse effect' (*i.e.*, the gravitational deflection of light) *refute* both E and N by yielding greater values than predicted by E and, *a fortiori*, greater values than those entailed by N. Incidentally, concerning N, Popper says there that 'N, even on Einstein's favourable interpretation, predicts at most half the result of E'. It is unclear why Popper thinks that the ratio between the deflection values predicted by N and E is only *at most* ½ and even then only on a favourable interpretation of N. Max Born ([1962], p. 359) has noted that if a ray of light is treated as, say, a comet which approaches the sun with the velocity of light, then—as J. von Soldner ([1804]) showed as early as 1801—N's theory of attraction predicts half the value yielded by E. And the exact value ½ of the ratio between the predicted deflection values is deduced by Tolman ([1934], pp. 210–11). Moreover, in his 1911 paper, Einstein ([1911], p. 108) himself first deduced only less than one half the deflection which he derived in 1916 from his full-blown E ([1916], p. 163).

As is evident from the *context* of the quotation in I, II above, Popper ([1972], p. 53 and n. 22) invokes Prop. P° and his thesis $QC(N) \subset QC(E)$ to claim that if $E_F \subseteq N_F$, then $N <_{vs} E$. I have already referred the reader to sections 2 and 3 of my ([1976*a*]) in this Journal for the arguments which prompt me to maintain that Prop. P° is at best ill-founded, and that the thesis $QC(N) \subset QC(E)$—which yields $Ct(N) \subset Ct(E)$—is false. Hence I now need to comment on the conclusion of Popper's aforecited assertion ([1972], p. 53) that if every known observational finding which refutes E likewise refutes N, then 'we have reason to suspect, or to conjecture' that $E_F \subseteq N_F$.[2]

If E is false, as assumed by Popper, then our result (IX) becomes applicable to the incompatible theories E and N, and it tells us that $E_F \nsubseteq N_F$! Thus our (IX) shows that even if every one of our tests which refutes E likewise refutes N, we can nevertheless *be sure* of the *falsity* of the conjecture $E_F \subseteq N_F$, which Popper wishes to countenance on the basis of putatively available refuting tests.

That this conjecture is *false* was unwittingly admitted by Popper himself *on empirical grounds* when he told us that the *numerically false prediction* made by E for the eclipse effect—I'll call this prediction 'p_e'—is *different* (at least *twice* as great) from the *false* corresponding prediction p_n made by N. Thus, by Popper's own account, $p_e \in E_F$ but $p_e \notin N_F$, so that $E_F \nsubseteq N_F$. More generally, the mere fact that an observationally false prediction made by E is *matched by* an observationally false prediction made by N, *in the sense* that a given observational finding (*e.g.* concerning

---

[1] We can easily give a very elementary proof that two logically incompatible *false* statements, say X and Y, cannot be rank-ordered with respect to $<_{vs}$. Since X and Y are each false while being logically incompatible, our pertinent weaker form of Tichý's result tells us at once that

$$Y \nless_{vs} X \text{ (here } X \text{ plays the role of our above false theory } B\text{)}$$

and

$$X \nless_{vs} Y \text{ (here } Y \text{ plays the role of our above false theory } B\text{).}$$

[2] When Popper says that 'the stronger theory [E] has no greater falsity content than its weaker predecessor [N]', he is thereby codifying the conjecture $E_F \subseteq N_F$ *rather than* merely *denying* that E_F *exceeds* N_F. For in this context, Popper is clearly assuming that E_F and N_F are *comparable* with respect to the subset relation. Without that assumption, the conjecture would not have any bearing on the deduction of greater verisimilitude via Prop. P°.

the eclipse effect) simultaneously shows both predictions to be false, ***hardly*** suffices to show that these two false predictions (*e.g.* p_e and p_n) ***are the same***. Hence unless these 'matching' false predictions ***are the same***, their matching in the stated sense cannot even serve to ***instantiate*** the conjecture $E_F \subseteq N_F$! In any case, we saw in the preceding paragraph that if E is false, then this conjecture must be false.

Hence if E is false as Popper assumes, then Popper is ***not*** even entitled to the premise $E_F \subseteq N_F$ as a basis for deducing $N <_{vs.} E$ via his unfounded Prop. P°.

But Popper's deduction of $N <_{vs.} E$ from $QC(N) \subset QC(E)$, $E_F \subseteq N_F$ ***and*** Prop. P° is not only ***ill-founded*** because of the gratuitousness or falsity of its premisses: We know from our results (VII) and/or (IX) as codified in our pertinent weaker form of Tichy's Prop. 2.4 that if E is false, then the conclusion of Popper's deduction is likewise false. Hence $N \nless_{vs.} E$.

Case (*c*)*:* A *is logically independent of a false theory* B

Clearly, this Case (*c*) is much less relevant to scientific growth of knowledge, if at all, than either Case (*a*) or Case (*b*).

There seem to be no particular features of this special case which are sufficient to exhibit in a more immediate and elementary way than Tichý's general proof that Tichý's Prop. 2.4 holds in this special case.

But consider the following Case (*d*), which is ***more general*** than Case (*c*) in regard to theory B, while being less general in regard to theory A: A ***true*** theory A is ***logically independent*** of a theory B which may be either true or false. Perhaps it is of interest to note that for this Case (*d*), an extremely simple proof can be given that $A \nless_{vs.} B$ as follows.

We are given the following three conditions:

(I) $A \in T$, so that $A \in A_T$
(II) $\sim(B \rightarrow A)$, so that $A \notin Cn(B)$
(III) $\sim (B \rightarrow \sim A)$

The conjunction of (II) and (III) asserts that A is logically independent of B.

From (I) and (II), we have

(IV) $A \in A_T$ but $A \notin B_T$.

But (IV) entails that

(V) $A_T \nsubseteq B_T$, whence it follows as in Case (*b*) that
(VI) $A \nless_{vs.} B$. Q.E.D.

This proof made no use of condition (III). Hence it applies to any A and B which satisfy conditions (I) and (II), regardless of whether they *also* satisfy (III) or not. A true A which is *logically independent* of B satisfies (I) and (II) as well as (III).

Case (*d*) is not covered by Tichý's Prop. 2.4.

Now recall that the antecedent of Prop. P includes the condition $B_F \subseteq A_F$, while the antecedent of Prop. $P°$ includes the condition $E_F \subseteq N_F$. Then I ask the following question:

(*ii*) *Can the Conjectures* $B_F \subseteq A_F$ *and* $E_F \subseteq N_F$ *be Warranted by Popperian Corroborations or Only Inductivistically, if At All?*

(*a*) *The Case of Compatible Competitors*

B unilaterally entails A. Here we are *given* that $Cn(A) \subset Cn(B)$, so that $A_F \subseteq B_F$. Hence $B_F \not\subset A_F$ and the conjecture that B_F is a subset of A_F would therefore be that $B_F = A_F$. Now note under what conditions Popper ([1972], p. 53) feels entitled to say in this case that 'we have reason to suspect, or to conjecture' that B_F is a subset of A_F. As will be recalled from our 1 (II) above, when speaking of the logically stronger of two theories (*i.e.*, our theory B here), he says: 'We try to find its weak points, to refute it. If we fail to refute it, or if the refutations we find are at the same time also refutations of the weaker theory [our theory A] which was its predecessor [n. 22 omitted here], then we have reason to suspect, or to conjecture, that the stronger theory has no greater falsity content than its weaker predecessor. . . '. I submit that the conditions which Popper here countenances as *reason* to conjecture that B_F is a subset of A_F qualify as such a reason only on the basis of a canon of evidential support which is the methodological *bête noire* of his entire philosophy: *induction by enumeration.*

For let us first suppose with Popper that *in our observational tests* of the predictions made by the stronger theory B, we find some to be false but such that all those B-predictions which turn out to be *false* are likewise predictions made by the theory A, and hence also belong to the falsity-content A_F of A. How then can this fact about the limited *sample* of B-predictions *which we have tested observationally* in regard to falsity be adduced as a *reason* for suspecting or conjecturing that *any and all false* consequences of B are likewise consequences of A and hence also belong to A_F? To offer the stated results as to the incidence of falsity in a *tested sample* of the consequences of B as a 'reason' for conjecturally generalising to the hypothesis that B_F is a subset of A_F is to invoke none other than *induction by enumeration* ('the straight rule').

It will be vital hereafter *not* to conflate putative corroborations or refutations of the stated conjectures concerning subset relations between specified falsity-contents, on the one hand, with corroborations or refutations of the competing theories B and A or E and N themselves, on the other. Thus, in our present case of ***compatible*** 'competitors' B and A, an observational finding which refutes both of them, by being contrary to one of their ***common*** observational consequences, clearly would ***not*** refute the conjecture $B_F = A_F$ but would exhibit an instance of the latter.

Since we are given that $A_F \subseteq B_F$, the only possible rival of the conjecture $B_F = A_F$ is $A_F \subset B_F$. Hence Popper's theory of corroboration would demand that if the conjecture $B_F = A_F$ is to be corroborated by successful predictions, the following conditions would have to be satisfied by the latter test outcomes: These outcomes would either have to refute the rival hypothesis $A_F \subset B_F$ or would have to be 'novel facts' with respect to the latter rival in the sense that they could at least *not* have been expected if the hypothesis $A_F \subset B_F$ is true. Clearly, the afore-stated putative observational findings do not refute the rival conjecture $A_F \subset B_F$. Nor do these putative observational findings seem to qualify as 'novel facts' in Popper's stated sense with respect to that rival conjecture.

Moreover, even if a supposed Popperian corroboration of the conjecture $B_F = A_F$ were *prima facie* feasible, in the case of a *false* B it would be just as unavailing as a *prima facie* inductivist confirmation of that conjecture: Under Case (*ai*) of subsection (*i*), we demonstrated both the *falsity* of $B_F = A_F$ {result (IV)} and the truth of $A_F \subset B_F$ {result (V)}.

Alternatively, consider Popper's other supposed outcome in which we *fail* to refute B. Then we have likewise failed to refute the logically weaker A. In this case as well, Popper takes these alternative findings to qualify as a ***reason*** for conjecturing that B_F is a (perhaps even empty) presumably improper subset of A_F. And indeed for all we know, under the alternative circumstances we just posited, A and even B *may* be true. Hence now we certainly cannot demonstrate the *falsity* of $B_F = A_F$. But by parity with our reasoning in the first case of our subsection (*a*) here, it is clear that Popper is (tacitly) appealing to induction by enumeration when he takes the alternative putative findings to qualify as a reason for conjecturing that $B_F = A_F$.

This charge of methodological inconsistency on his part is not gainsaid or mitigated by Popper's disclaimer ([1972], p. 58) that 'My defence of the legitimacy of the idea of verisimilitude has sometimes been grossly misunderstood. In order to avoid these misunderstandings it is advisable to keep in mind my view that not only are all theories conjectural, but also all appraisals of theories, including comparisons of theories from the point of view of their verisimilitude'. This disclaimer cannot serve to circumvent Popper's appeal to enumerative induction, since by his own canons, merely positive but *non*-corroborating instances of a conjecture do not qualify as a 'reason' for it: As Popper ([1962], p. 36) told us with attractive enthusiasm, 'Confirmations should count only if they are the result of *risky predictions*; that is to say, if, unenlightened by the theory in question, we should have expected an event which was incompatible with the theory—an event which would have refuted the theory'. (Cf. 2 above).

(*b*) *The Case of Incompatible Competitors:* E *and* N

Since E and N are incompatible, their Tarksian consequence classes are merely overlapping rather than being comparable with respect to the subset or superset relation. And since at least one of them—E or N—must be false while not belonging to the other's consequence class, we know that $E_F \neq N_F$. Furthermore, the very findings which prompted the rejection of N in favor of E by physicists can reasonably be taken to have *refuted* the hypothesis $N_F \subset E_F$. But since we have no guarantee that the falsity-contents E_F and N_F are at all comparable with respect to the subset or superset relation, the known falsity of $N_F \subseteq E_F$ does *not* enable us to deduce $E_F \subset N_F$. Hence we are left with both of the following rival hypotheses: (*i*) E_F and N_F are *incomparable* with respect to set inclusion, and (*ii*) $E_F \subset N_F$, where it is to be noted that we have *not yet* taken account of the following empirical findings noted by Popper ([1972], p. 53, *n.* 22): The mutually *contrary* predictions p_e and p_n of gravitational deflection of light made by E and N respectively are each observationally false, so that E is false no less than N. Clearly, as we already noted under Case (*b*) in section 3(*b*)(*i*), the fact that $p_e \in E_F$ while $p_e \notin N_F$ exhibits the falsity of the conjecture $E_F \subseteq N_F$, and yet Popper tries to base this conjecture upon just these 'eclipse effect' observations reported by him. Besides, even *prima facie* these observational findings do *not* pass muster as Popperian corroborations of the latter conjecture *vis-à-vis* the rival hypothesis of set-theoretical incomparability of E_F and N_F.

But since Popper nonetheless takes the gravitational deflection observations as 'reasons' for conjecturing that $E_F \subseteq N_F$, one can surmise only that he is tacitly appealing to enumerative induction, unmindful of the empirically known falsity of this conjecture.

(*iii*) *Conclusion*

It remains for the next section 3(*c*) to deal with the role of Prop. $P\#$ *vis-à-vis* Popper's *quantitative* verisimilitude. But from my analysis so far, I conclude that just as severe tests are *deductivistically* no more promising than non-severe ones in regard to weeding out false theories, so also Popper's thesis of the epistemic superiority of severe tests cannot be vindicated by his theory of *qualitative* verisimilitude. And my reasons are as follows:

(*a*) In the scientifically typical case of logically incompatible competitors, this vindication is an abortive non-starter, because the foundational Prop. P° as well as its antecedent condition $Ct(N) \subset Ct(E)$ are at best *ill-founded, even if* E *is actually true while* N *is false.* And if E is false, the verisimilitude-ranking in the consequent of Prop. P° is false, if only because of the falsity of the necessary condition $E_F \subseteq N_F$. Besides, even *prima facie* the presumed empirical findings which Popper adduces as 'reason' for the conjecture $E_F \subseteq N_F$ qualify as such at best only inductivistically, and are patently non-corroborating by Popper's own requirements.

(*b*) In the scientifically atypical case of *compatible* 'competitors' A and B, Prop. P is well anchored in Popper's theorem on truth-content. But, when the stronger theory B is false, Prop. P is *unavailing* as a vindication of severe tests of B, as already pointed out by Miller. And I add that even when both A and B are true, Popper's positive observational results from severe tests of B do *not* qualify as a 'reason' for conjecturing that the condition $B_F \subseteq A_F$ in the antecedent of Prop. P is satisfied, unless one countenances induction by enumeration (the straight rule). For the observational results in question surely fail to pass muster as Popperian corroborations of this conjecture.

(*c*) *Popper's* Quantitative *Theory of the Content and Verisimilitude of a Hypothesis, Prop.* P#, *and the Merits of Severe Tests*

Clearly a major aim of Popper's *quantitative* theory of the content and verisimilitude of hypotheses was to achieve comparability of any two theories, even if only in principle rather than in practice ([1972], pp. 58–9). It would be a major achievement of his quantitative theory if it could do the following: without creating new difficulties of its own, it would overcome the stated defects of the *qualitative* version for just those pairs of hypotheses A and B in which A and B address themselves to the same problems: Incompatible competitors—our Case (*b*) in 3(*b*)(*i*) *above*—and compatible 'competitors', which were our Case (*ai*) there. And such an achievement would be considerable, even if the quantitative theory could not also serve to vindicate severe tests either by itself or via some such proposition as $P\#$.

I shall comment on Popper's quantitative theory of content and verisimilitude as such and will deal with its bearing, if any, on the vindication of severe tests. It will be useful to proceed in the following order: (*i*)

First, I shall give an essentially formal comparison of the similarities and differences between Popper's quantitative verisimilitude *Vs.* and his qualitative *vs.* while *deferring* comment on his *probabilistic definitions* of his *measures* of content, truth-content and falsity-content; (*ii*) next, I shall comment critically on these probabilistic measures; and (*iii*) finally, I shall consider the bearing, if any, of Popper's quantitative theory of content and *Vs.* on the vindication of his advocacy of severe tests.

(*i*) *Comparison of Vs with vs.*

Popper denotes his *measure* of the content of a theory A by '$ct(A)$', and his respective *measures* of the truth-content A_T and falsity-content A_F of A by '$ct_T(A)$' and '$ct_F(A)$'. Of the two *quantitative* definitions of verisimilitude given by Popper ([1962], p. 234 and 396–7, and [1972], p. 334) in terms of ct_T and ct_F, we need to consider only the simpler of the two for our purposes. This definition ([1962], p. 234) is

$$\text{(0)}\quad Vs(A) =_{\text{Def.}} ct_T(A) - ct_F(A).$$

Note that we have used '*vs.*' to denote *qualitative verisimilitude* while using '*Vs*' to denote *quantitative* verisimilitude.

Now consider two theories A and B. According to Popper's ([1962], p. 394, item (vii)), the respective pairs of values $ct_T(A)$, $ct_F(A)$ and $ct_T(B)$, $ct_F(B)$ which determine $Vs(A)$ and $Vs(B)$ respectively via (0) above must satisfy the respective relations

$$\text{(1)}\quad ct_T(A)+ct_F(A) \geqslant ct(A)$$

$$\text{(2)}\quad ct_T(B)+ct_F(B) \geqslant ct(B).$$

We are interested in theories B and A such that both

$$\text{(3)}\quad ct(B) > ct(A)$$

and

$$\text{(4)}\quad Vs(B) > Vs(A).$$

Using (0) above, we can write

$$\text{(5)}\quad Vs(B) = ct_T(B) - ct_F(B)$$

and

$$\text{(6)}\quad Vs(A) = ct_T(A) - ct_F(A)$$

Subtracting (6) from (5), we get

$$\text{(7)}\quad Vs(B) - Vs(A) = ct_T(B) - ct_T(A) - [ct_F(B) - ct_F(A)]$$

The left-hand side of (7) is the *increment* in *Vs.* from A to B. We shall abbreviate the latter by writing

$$\text{(8)}\quad Vs(B) - Vs(A) =_{\text{Def.}} \Delta_A^B\, Vs.$$

We shall also use the additional abbreviations

(9) $ct_T(B) - ct_T(A) =_{\text{Def.}} \Delta^B_A\, ct_T$

and

(10) $ct_F(B) - ct_F(A) =_{\text{Def.}} \Delta^B_A\, ct_F$.

Using (8), (9) and (10), we can rewrite (7) as

(11) $\Delta^B_A\, Vs = \Delta^B_A\, ct_T - \Delta^B_A\, ct_F$.

Clearly, the rank-ordering of theories with respect to *qualitative* verisimilitude, if feasible at all, is restricted to those relatively rare pairs of theories whose respective truth-contents and falsity-contents sustain appropriate subset relations.

Popper thinks that, at least in principle, this deficiency of the *qualitative* theory of *vs.* is remedied by his *quantitative Vs.* For Popper claims ([1962], p. 234) that for any two theories A and B, the measures $ct(A)$, $ct(B)$, $ct_T(A)$, $ct_F(A)$, $ct_T(B)$ and $ct_F(B)$ are each objectively determined by specified logical probabilities, whether we humans happen to know them or not. And *if* there is indeed such objective determinateness, then *any two* theories A and B become comparable with respect to both *content* and *verisimilitude*, because the stated measures will then be *arithmetically* comparable not only in regard to more or less but also in regard to equality.

We shall shortly be contrasting qualitative *vs.* with quantitative *Vs.* more specifically after first remarking on a similarity between them. We are concerned with theories B and A such that

(12) $\Delta^B_A\, Vs > 0$.

Let us now set side by side the sufficient *and* necessary conditions for an increase in *quantitative* verisimilitude in the sense of (12), on the one hand, with the sufficient and necessary conditions for a corresponding excess Δ^B_A *vs.* in *qualitative* verisimilitude, on the other. In this collation, the arithmetical relations $<$ and $\leqslant$ will be the respective *counterparts* of the proper subset relation $\subset$ and the subset relation $\subseteq$.

It can be seen from Popper's definition of excess Δ^B_A *vs.* ([1962], p. 233) as rendered by the Tichý formulation given in our 3(*b*)(*i*) above that the following disjunction is the *quantitative counterpart* of the *necessary* and *sufficient* condition for excess Δ^B_A *vs.*, *i.e.*, for $A <_{vs.} B$:

Either

$$ct_T(A) < ct_T(B) \textit{ and } ct_F(B) \leqslant ct_F(A)$$

or

$$ct_T(A) \leqslant ct_T(B) \textit{ and } ct_F(B) < ct_F(A).$$

Popper notes ([1962], p. 234) that his definitions of excess Δ^B_A *vs.* and of Δ^B_A *Vs.* are such that the ***quantitative counterpart*** of the ***sufficient*** condition for excess Δ^B_A *vs.* is indeed also a ***sufficient*** condition for $\Delta^B_A\ Vs > 0$. This fact will soon become evident upon comparing the quantitative counterpart of the sufficient condition for $A <_{vs.} B$ with our (7) or (11) above. But we shall then see that the definitions of Δ^B_A *Vs.* and of excess Δ^B_A *vs.* are also such that the ***quantitative counterpart*** of the necessary condition for excess Δ^B_A *vs.* is ***not*** also a ***necessary*** condition for $\Delta^B_A\ Vs > 0$!

Now recall (11), which asserts $\Delta^B_A\ Vs = \Delta^B_A\ ct_T - \Delta^B_A\ ct_F$. It is then clear that the necessary and sufficient condition for

$$\Delta^B_A\ Vs > 0$$

is that

(13) $\Delta^B_A\ ct_T > \Delta^B_A\ ct_F$.

Using the notation of (9) and (10) above, the stated disjunctive quantitative counterpart of the necessary and sufficient condition for excess Δ^B_A *vs.* (*i.e.*, $A <_{vs.} B$) becomes:

Either

$$\Delta^B_A\ ct_T > 0 \textit{ and } \Delta^B_A\ ct_F \leqslant 0$$

or

$$\Delta^B_A\ ct_T \geqslant 0 \textit{ and } \Delta^B_A\ ct_F < 0.$$

Let us call the latter disjunction 'Condition *C*'. Clearly Condition *C* satisfies the ***sufficient*** condition (13) for $\Delta^B_A\ Vs > 0$.

But let us now see why Condition *C* is ***not*** necessary for $\Delta^B_A\ Vs > 0$, although it is the quantified counterpart of the necessary condition for $A <_{vs.} B$. Note first that ***each*** disjunct in Condition *C* requires at least ***both of the following***:

(*i*) $\Delta^B_A\ ct_T \geqslant 0$, and (*ii*) $\Delta^B_A\ ct_F \leqslant 0$.

Hence Condition *C* requires both (*i*) and (*ii*). But we shall now see that neither (*i*) nor (*ii*) is required by $\Delta^B_A\ Vs > 0$. It will then follow that $\Delta^B_A\ Vs > 0$ does ***not*** require Condition *C* as being necessary. (For if it did, it would also have to require whatever is required by Condition *C*—such as (*i*) and (*ii*)—which we shall now see it does not.)

(1) That (*i*) $\Delta^B_A\ ct_T \geqslant 0$ is ***not*** required by an increase in quantitative *Vs* (*i.e.*, by $\Delta^B_A\ Vs > 0$) is clear from the fact that the latter ***can*** be satisfied via (13) by a ***negative*** value of $\Delta^B_A\ ct_T$ along with a ***lesser*** negative value of $\Delta^B_A\ ct_F$. Intuitively, the ***drop*** in truth-content from *A* to *B* is more than compensated by the ***greater drop*** in falsity-content, thus making for *B*'s ***greater*** *Vs*! Incidentally, if the latter is the case for theories *A* and *B* such that $ct(B) > ct(A)$, then the truth-content and falsity-content values which generate these ***negative increments*** $\Delta^B_A\ ct_T$ and $\Delta^B_A\ ct_F$ via equations (9) and (10) must also meet the following demand: The values of $ct_T(A)$,

$ct_F(A)$, $ct_T(B)$ and $ct_F(B)$ in question must satisfy the relations (1) and (2) *above* so as to *allow* that the right-hand sides of (1) and (2) fulfil the condition $ct(B) > ct(A)$.

(2) That (*ii*) $\Delta_A^B\, ct_F \leqslant 0$ is *not* required by a positive value of $\Delta_A^B\, Vs$ is clear from the following fact: $\Delta_A^B\, Vs > 0$ can be achieved via (13) by a *positive* value of $\Delta_A^B\, ct_F$ along with a *greater* positive value of $\Delta_A^B\, ct_T$. Intuitively, the falsity content may increase, so long as the truth-content increases *even more*. Incidentally, if the latter is the case when $ct(B) > ct(A)$, then remarks corresponding to those just made under (1) *above* apply.

Thus, unlike Condition C, $\Delta_A^B\, Vs > 0$ requires neither that $\Delta_A^B\, ct_T$ be *non*-negative nor that $\Delta_A^B\, ct_F$ be *non*-positive. Hence we have shown that Condition C is not necessary for an increase in quantitative verisimilitude.

But what is the import of the fact that the requirements for greater *quantitative* verisimilitude are much *less stringent*—in this specific sense as well as in the stated general respect of purported unrestricted comparability—than the requirements for an excess of *qualitative* verisimilitude? In (*iii*) below, we shall consider whether this lesser stringency enables the *quantitative* theory to succeed where the qualitative one failed in vindicating severe tests within the framework of Popper's deductivistic theory of corroboration.

(*ii*) *Popper's Probabilistic Measure of Content, and the Bearing of his Quantitative Concept of Content on his* Reductio ad Absurdum *Argument against Probabilistic Inductivism.*

As will be recalled from our 1(II) *above*, Popper ([1972], p. 53) begins his explanation of how he conceives Prop. P to be 'the logical basis of the method of science' by saying: 'A theory is the bolder the greater its content. It is also the riskier: it is the more probable to start with that it will be false'. First I wish to inquire whether the latter assertion of greater riskiness has the sanction of the probability calculus and to point out how this assertion fails to cohere with one of Popper's major arguments against the notion of inductive probability.

Suppose that B entails A unilaterally. Then there clearly is warrant for saying that in virtue of exceeding A in content, B is also 'bolder' than A: While *every* observational finding that refutes the logically weaker theory also refutes the stronger one, the converse does not hold. For there are some logically possible findings which would refute the stronger theory without refuting the weaker one. So much for boldness and greater content.

But Popper goes on to say that the theory of greater content 'is also the riskier: it is the more probable to start with that it will be false'. And he had sought to justify this statement by his earlier claim ([1972], p. 51) that 'the more information a statement carries, the smaller will be the logical [*i.e.*, *a priori* mathematical] probability that it is true (accidentally, as it were)'. Hence let us see whether the mathematical probability calculus sanctions the following two assertions made by Popper ([1962], p. 218; *cf.* also p. 286): '... with increasing [Tarskian] content, probability decreases, and *vice versa*', and 'if growth or [of] knowledge means that we operate with theories of increasing content, it must also mean that we operate with theories of decreasing probability (in the sense of the calculus of probability)'.

Popper is avowedly concerned with universally quantified hypotheses B and A whose quantifiers range over a domain of infinitely many individuals. But at least for such 'infinitistic' universal hypotheses, the pertinent theorem of the calculus of probabilities is as follows (Carnap [1967], sections 58 and 59[1]): If B *unilaterally* entails A, then the logical probability $p(B)$ of B is less than *or equal to* the logical probability $p(A)$ of A. Contrary to Popper, at least for the specified kind of universal hypotheses, the mathematical probability calculus does *not* entail his cited *stronger* assertion that $Cn(A) \subset Cn(B) \rightarrow p(B) < p(A)$.

Indeed, as I have documented elsewhere in greater detail (Grünbaum [1976*b*], section 3(I), pp. 39–40), Popper ([1959], pp. 363–6, and [1962], pp. 192–3) flatly contradicts his own stronger assertion of *lesser* probability by offering the following claim as a *reductio ad absurdum* of the concept of inductive probability: Both the *logical* probability and the *a posteriori* probability of *any* universal law hypothesis are zero, and hence 'all [universal] theories, including the best, have the same probability, namely zero' (Popper [1962], p. 192). But the latter thesis commits him to opt for the *equality* of $p(B)$ and $p(A)$ even when B unilaterally entails A, thereby contradicting his other assertion that in this very case, $p(B)$ must be *less than* $p(A)$! Moreover, since Popper defines the measure $ct(T)$ of the content of a theory T as $1-p(T)$, his one claim that all universal theories have *zero* logical probability has the following consequence: Even when a universal B *unilaterally* entails A, $ct(B) = ct(A)$. Suppose that for the sake of consistency, Popper were to *retract* his thesis that for any 'infinitistic' universal theory T, $p(T) = 0$. Then he could still not adduce against the concept of inductive probability that '*the more a statement asserts, the less probable it is*' (Popper [1962], p. 286; italics in original). For, as we noted, the probability calculus entitles him to only the weaker assertion that if $Cn(A) \subset Cn(B)$, then $p(B)$ is less than *or equal to* $p(A)$. So much for the serious inconsistency between Popper's account of the relation of logical probability to Tarskian content, on the one hand, and his *reductio ad absurdum* attempt to establish 'the impossibility of an inductive probability' ([1959], p. 363), on the other.

---

[1] I am indebted to Teddy Seidenfeld for this reference and for clarifying comments on the conditions of validity for the theorem in question.

Quite apart from this contradiction, the *failure* of the probability calculus to *guarantee* Popper's assertion

$$Cn(A) \subset Cn(B) \rightarrow p(B) < p(A)$$

has the following unfortunate consequence for his proposed measure $ct\ (T)$ of the content of a theory T: Since he defines ([1962], pp. 390–2; [1972], p. 51)

$$ct(T) =_{\text{Def.}} 1-p(T),$$

the probability calculus does *not* entitle him to his contention that ([1972], p. 51)

$$Cn(A) \subset Cn(B) \rightarrow ct(B) > ct(A);$$

instead, the probability calculus vouchsafes only the *weaker* assertion

$$Cn(A) \subset Cn(B) \rightarrow ct(B) \geqq ct(A).$$

In other words, Popper's probabilistic measure $ct(T)$ is such that when B *unilaterally* entails A, *there is no guarantee* that $ct(B) > ct(A)$, since the probability calculus does *not* rule out that $ct(B) = ct(A)$! And yet Popper ([1972], p. 51) had told us the following:

> As I have shown in various places, the content A of a will be the greater, the smaller the logical probability $p(a)$ or $p(A)$. For the more information a statement carries, the smaller will be the logical probability that it is true (accidentally, as it were). We can therefore introduce a '*measure*' of the content (it can be used mainly topologically, that is, as an indicator of linear order),
>
> $$ct(a),$$
>
> that is, the (absolute) content of a. . . . These 'measures' ct can be defined with the help of the calculus of probability . . .

I find it hard even to speculate how Popper thought he could justify his pretentious claim that in the case of the *incompatible* theories E and N, $ct(N) < ct(E)$ ([1972], p. 53): He just *tells* us that the intuitive judgment $QC(N) \subset QC(E)$ 'can be shown' to be 'borne out by the content measures $ct(N)$ and $ct(E)$'!

Note that he had given the following probabilistic definitions of his *measures* $ct_T(A)$ and $ct_F(A)$ of truth-content and falsity-content:

$$ct_T(A) =_{\text{Def.}} 1-p(A_T)$$
$$ct_F(A) =_{\text{Def.}} 1-p(A_F).$$

In a more cautious or less sanguine frame of mind, Popper ([1962], pp. 233–4) had characterised the 'assumption that the content and truth-content of a theory *a* are in principle *measurable*' as 'perhaps fictitious'. Let me briefly elaborate on this possibility of fictitiousness in its bearing on the following thesis of his ([1972], pp. 51, 52 n. 20, and 59): In principle, though normally not known to us, the measures $ct(B)$ and $ct(A)$ are *determinate* at least up to assuring a linear order such that

$$Cn(A) \subset Cn(B) \rightarrow ct(B) > ct(A).$$

As Colin Howson has pointed out to me, several doubts have been raised concerning this claim of determinateness:

1. At least before Tichý's [1976], in all theories in which *prima facie* determinate logical probability measures are assigned to statements, these measures are indeterminate in the quite strong sense of being language-dependent, so that no determinate measure is assignable to an arbitrary theory.

2. Generalising a result of Carnap's, it can be shown that for any first order language L, and for any (non-tautological and non-contradictory) sentence σ of L, and any r in the interval $0 \leqslant r \leqslant 1$, there is a probability measure p, such that p assigns r to σ.[1]

In Tichý's [1974], he takes the logical probability of a statement A to be the ratio of the cardinality of the range A—*i.e.*, of the class of states of affairs for which A is true—to the cardinality of the entire logical space. At least *prima facie*, in the case of universal 'infinitistic' theories like E and N for which the universe of discourse and the logical space are infinite, the difficulty of indeterminateness of the logical probabilities then presents itself. Having been aware of this *prima facie* difficulty, Tichý ([1976]) has offered an alternative to Popper's theory of *Vs.* in which Tichý adapted an idea of Hintikka's so as to exhibit a 'natural' way in which the indeterminateness of the logical probabilities can be overcome. It remains to be seen whether Tichý's proposal for achieving determinateness will bear further scrutiny in the face of doubts raised by David Miller and John Worrall.[2] If it does, then Tichý will have rescued the logical probabilities needed by Popper's quantitative theory of content and *Vs.* from the fictitiousness of indeterminateness which otherwise besets them.

(*iii*) *Can Popper's Quantitative Theory of Content and* Vs. *Give a Deductivistic Vindication of his Advocacy of Severe Tests?*

Popper ([1972], pp. 58–9) issued the following clarifying disclaimer:

---

[1] For a sketch of Howson's own view of the logical interpretation of probability, see his ([1975], section 3, pp. 194–8).

[2] I am indebted to Pavel Tichý, John Worrall, David Miller and Colin Howson for correspondence on this point.

> Why, then, have I tried to show that verisimilitude can be defined, or reduced to, other terms (truth content, falsity content and, in the last instance, logical probability)?
>
> Some people have assumed that my aim was something like exactness or precision; or even applicability: that I hoped to find a numerical function which can be applied to theories and which tells us, in numerical terms, what their verisimilitude is (or at least their truth content; or perhaps their degree of corroboration).
>
> In fact, nothing can be further removed from my aims. I do not think that degrees of verisimilitude, or a measure of truth content, or falsity content (or, say, degree of corroboration, or even of logical probability) can ever be numerically determined, except in certain limiting cases (such as 0 and 1). And although the introduction of a measure function makes all contents comparable in principle, or in theory, I believe that in actual application we depend entirely on those rare cases which are comparable on non-metrical and, as it were, qualitative or general logical grounds, such as cases of logically stronger and weaker *competing* theories; that is, theories aimed at solving the same problems. For actual comparison we depend entirely upon these cases (paradoxically, one might say, since measure functions such as probabilities make their arguments in principle *generally* comparable).
>
> What then, one might ask, is the point of my attempts to show that verisimilitude is definable in terms of logical probability? My aim is to achieve (on a lower level of precision) for verisimilitude something similar to what Tarski achieved for truth: the rehabilitation of a commonsense notion which has become suspect, but which is in my opinion much needed for any critical commonsense realism and for any critical theory of science.

In this subsection I wish to exhibit some of the specific details which lend substance to Popper's *disclaimer* of *applicability* for the quantitative concept *Vs.* by particular reference to the case of the deductivistic vindication of severe tests.

As we saw in subsection (*i*) of 3(*c*), item (13), the necessary and sufficient condition for $\Delta_A^B\ Vs > 0$ is that

$$\Delta_A^B\ ct_T > \Delta_A^B\ ct_F.$$

For greater notational simplicity, I shall hereafter denote the truth-content increment $\Delta_A^B\ ct_T$ by 'Δ_T' and the falsity-content increment $\Delta_A^B\ ct_F$ by 'Δ_F', while letting 'ΔVs' be short for '$\Delta_A^B\ Vs$'.

In contrast to the counterpart requirements for excess *qualitative vs.*, the fulfilment of the desideratum $\Delta Vs > 0$ neither requires Δ_T to be non-negative nor depends on Δ_F's having a non-positive value: Clearly, both increments might be negative or both might be positive and yet could yield a positive value for ΔVs.

It might be thought, however, that whenever B unilaterally entails A, we could deduce quite generally from Popper's theorem on truth-content ([1966], p. 350) that $\Delta_T > 0$, so that the fulfilment of $\Delta Vs > 0$ would then depend only on the value of the falsity-content increment Δ_F. And if that were so, the falsity-contents of B and A would thus play a decisive role after all, even when a *false* B unilaterally entails A, unencumbered by the fact that Prop. P is then rendered unavailing by the qualitative relation $A_F \subset B_F$.

To see why this supposition seems to be doubtful, first note the statement of Popper's *qualitative* theorem on truth-content ([1966], p. 350; [1972], p. 52, n. 20):

$$[Cn(A) \subset Cn(B)] \longleftrightarrow A_T \subset B_T$$

And recall that $\Delta_T = p(A_T) - p(B_T)$. Then how is it *guaranteed* that $(A_T \subset B_T) \rightarrow \Delta_T > 0$? If the logical probabilities of the respective *sets* A_T and B_T were such that the latter entailment is vouchsafed by the probability calculus, then $\Delta_T > 0$ would, of course, follow from $Cn(A) \subset Cn(B)$ via the theorem on truth-content. Furthermore, $\Delta_T > 0$ could not be deduced when B unilaterally entails A even if we assume *merely for the sake of argument* that Popper could prove the following quantitative *counterpart* of his theorem on truth-content:

$$[ct(A) < ct(B)] \longleftrightarrow [\Delta_T > 0]$$

Even then the latter premiss would *not* permit the deduction of $\Delta_T > 0$ from $Cn(A) \subset Cn(B)$. For, as I was at pains to show in the preceding subsection (*ii*), the required missing premiss $Cn(A) \subset Cn(B) \rightarrow ct(A) < ct(B)$ is gratuitous.

These results enable us to make a comment on the status of Prop. $P\#$. Although $\Delta_T > 0$ and $\Delta_F \leqslant 0$ is a sufficient condition for $\Delta Vs > 0$, the following statement holds: When B unilaterally entails A, Popper's qualitative theorem on truth-content cannot be claimed to guarantee the deducibility of the quantitative version $P\#$ of Prop. P, at least as long as it is not shown that $(A_T \subset B_T) \rightarrow \Delta_T > 0$. Moreover, the premiss $ct(A) < ct(B) \rightarrow Cn(A) \subset Cn(B)$—which is equivalent to the premiss $p(A) > p(B) \rightarrow Cn(A) \subset Cn(B)$—cannot be invoked in the putative deduction. And even if Prop. $P\#$ were validly deducible, it would be unavailing, since B's unilateral entailment of A does not guarantee that the condition $ct(A) < ct(B)$ in the antecedent of Prop. $P\#$ will be satisfied.

In the case of logically *incompatible* theories like E and N, Popper's qualitative theorem on truth-content is plainly useless as a premiss for attempting to derive $\Delta_T > 0$, and the invocation of its stated quantitative counterpart is presumably gratuitous. By the same token, the assertion of Prop. $P\#$ is presumably gratuitous in this case.

Thus it would seem that for compatible competitors no less than for incompatible ones, Popper's *quantitative* theory of content and $Vs.$ fails to confer a crucial role on Δ_F in regard to determining whether $\Delta Vs > 0$ or not.

Could we perhaps corroborate $\Delta Vs > 0$ directly? The hypothesis $\Delta Vs > 0$, which asserts that $\Delta_T > \Delta_F$, would be false if either $\Delta_T = \Delta_F$ or $\Delta_T < \Delta_F$. Could putative successful outcomes of severe tests *of the theory B* qualify as Popperian corroborations of the hypothesis $\Delta Vs > 0$ *vis-à-vis* either of its two rivals $\Delta_T = \Delta_F$ and $\Delta_T < \Delta_F$? In particular, could at least mostly successful severe tests of B corroborate $\Delta Vs > 0$ in the case of a *false B*, even though $A \nless_{vs.} B$ in that case? Suppose that Tichý's [1976] can be presumed to have solved the problem of the *prima facie* objective indeterminateness of the value of Δ_T—*i.e.*, $p(A_T) - p(B_T)$—and of Δ_F, *i.e.*, $p(A_F) - p(B_F)$, to within the kind of rank-ordering of Δ_T and Δ_F needed here. It is then still at best very unclear how putative outcomes of severe tests of B could possibly serve to furnish *Popperian corroborations* of $\Delta Vs > 0$ *vis-à-vis* its stated rivals. (Recall here our discussion in 3(*b*)(*ii*) of the much less obscure attempt to corroborate the *qualitative counterpart* of the hypothesis $\Delta_F \leqslant 0$.) Clearly, we must not conflate here corroborations of the theory B with corroborations of the hypothesis $\Delta Vs > 0$.

Indeed, to my mind it is not even clear how an *enumerative inductivist* could adduce the incidence of success and failure in severe tests of B as a basis for making a *comparative* 'estimate' concerning the values of Δ_T and Δ_F. Lakatos ([1968], pp. 393–9) seems to claim (p. 397) that meaningful 'estimates' of ΔVs. are possible when the corroborations of the theories A and B are comparable. If that should be so, then my discussion of the hypothesis $\Delta_T > \Delta_F$ is consonant with the following *caveat* issued by Lakatos ([1968], p. 397):

> Even where comparisons [of the corroboration of two theories] are possible, one can easily conceive of conditions which would make the estimate of verisimilitude by corroboration false. The successive scientific theories may be such that each increase of truth-content could be coupled with an even larger increase in hidden falsity-content, so that the growth of science would be characterised by increasing corroboration and decreasing verisimilitude.

Lakatos then illustrates this risk by calling attention to the epistemic limitations arising from the cosmic parochialism of our small spatio-temporal 'corner' in the universe.

As we noted, Popper himself ([1972], p. 58) pointed out the latter possibility as a risk inherent in comparative conjectural appraisals of theories with respect to verisimilitude. And I trust that I have exhibited the justification for his *disclaimer* of the *epistemic applicability* of his *quantitative* theory of *Vs.* by reference to the quest for a deductivistic vindication of severe tests.

But this leaves us with the overall conclusion which I stated anticipatorily in 1 above: Popper has not given us a deductivistically viable rationale for his methodological advocacy of severe tests as epistemically more promising than non-severe tests. And thus one wonders whether any kind of deductivistic rationale can be given for *normatively* espousing the method of bold conjectures *and attempted refutations* as the method of science.

University of Pittsburgh

REFERENCES

BORN, M. [1962]: *Einstein's Theory of Relativity*. Revised edition.

CARNAP, R. [1967]: *Logical Foundations of Probability*. 2nd edition.

EINSTEIN, A. [1911]: 'On the influence of gravitation on the propagation of light', in *The Principle of Relativity*, A Collection of Original Memoirs, 1952, pp. 99–108.

EINSTEIN, A. [1916]: 'The foundation of the general theory of relativity', in *The Principle of Relativity*, A Collection of Original Memoirs, 1952, pp. 111–64.

GRÜNBAUM, A. [1976*a*]: 'Can a theory answer more questions than one of its rivals?', *The British Journal for the Philosophy of Science*, **27**, pp. 1–23.

GRÜNBAUM, A. [1976*b*]: 'Is falsifiability the touchstone of scientific rationality? Karl Popper versus inductivism', in R. S. Cohen and M. Wartofsky (*eds.*): Lakatos Memorial Volume, *Boston Studies in the Philosophy of Science*, **39**.

HARRIS, J. H. [1974]: 'Popper's definitions of verisimilitude', *The British Journal for the Philosophy of Science*, **25**, pp. 160–6.

HOWSON, C. [1973]: 'Must the logical probability of laws be zero?', *The British Journal for the Philosophy of Science*, **24**, pp. 153–60.

HOWSON, C. [1975]: 'The rule of succession, inductive logic, and probability logic', *The British Journal for the Philosophy of Science*, **26**, pp. 187–98.

LAKATOS, I. [1968]: 'Changes in the problem of inductive logic', in I. Lakatos (*ed.*): *The Problem of Inductive Logic*, pp. 393–7.

MILLER, D. [1974]: 'Popper's qualitative theory of verisimilitude', *The British Journal for the Philosophy of Science*, **25**, 166–77.

POPPER, K. [1959]: *The Logic of Scientific Discovery*.

POPPER, K. [1962]: *Conjectures and Refutations*.

POPPER, K. (1966]: 'A theorem on truth-content', in P. K. Feyerabend and G. Maxwell (*eds.*): *Mind, Matter, and Method*, pp. 343–53.

POPPER, K. [1972]: *Objective Knowledge*.

SALMON, W. [1966]: *The Foundations of Scientific Inference*.

SOLDNER, J. von [1804]: *Bode Astron. Jahrb.*, pp. 161–72.

TICHÝ, P. [1974]: 'On Popper's definitions of verisimilitude', *The British Journal for the Philosophy of Science*, **25**, pp. 155–60.

TICHÝ, P. [1976]: 'Verisimilitude redefined', *The British Journal for the Philosophy of Science*, **27**, pp. 25–42.

TOLMAN, R. C. [1934]: *Relativity, Thermodynamics and Cosmology*.

Brit. J. Phil. Sci. 27 (1976), 137–146 *Printed in Great Britain*

Ideal Types and Empirical Theories

by DAVID PAPINEAU

How can ideal types be of any help in understanding empirical reality?

The present paper is addressed to this question. In the first two sections I shall explain why it needs answering and clear away some of the basic confusions which have bedevilled discussion of this issue. In the following two sections I shall criticise two of the standard approaches to the subject. Finally I suggest an answer.

1 I shall take it that a distinguishing characteristic of ideal type concepts is that they have no instances. That is, ideal type concepts differ from other descriptive concepts in that there are no actual situations, entities, events, phenomena, or whatever, which satisfy them. (In general, however, there will be cases which in some intuitive sense 'approximate' to any ideal type concept—this will be of some significance later.)

Examples of ideal type concepts are thus the sociologist's notion of a 'pure rational bureaucracy', the economist's idea of a 'perfectly competitive market' or the physicist's idea of a 'perfect (terrestrial) vacuum'.

The reason why the use of ideal type concepts requires some explanation or justification is obvious enough: unlike other scientific concepts, ideal type concepts cannot be used in the formulation of testable generalisations whose acceptance enables the explanation and prediction of actually occurring phenomena. For suppose we accept a generalisation with an ideal type concept as antecedent term specifying its range of application, such as, say, 'All pure rational bureaucracies are maximally efficient at achieving organisational goals', or 'All bodies falling in a perfect vacuum have constant acceleration'. Since ideal types have no instances, accepting such a generalisation will never allow us to infer from the fact that one state of affairs obtained, or obtains, the further fact that something else then had to, or will, happen—we will never actually come across any state of affairs to which we might apply the generalisation. Generalisations with ideal type concepts as consequent terms are obviously no good for explanation or prediction either, since the things they would allow us to explain or predict never happen. (Though there is clearly a distinction between generalisations with ideal type concepts as antecedent terms and

Received 5 June 1975

those with ideal concepts as consequent terms, the use of these two types of generalisations raises similar problems and can be accounted for along similar lines. For convenience, however, I shall restrict my discussion to the former type of generalisation. Henceforth I shall use the expression 'ideal type generalisation' to refer to this kind of generalisation only.)

In response to this *prima facie* argument for the uselessness of ideal type generalisations some writers have sought to defend the adoption of ideal type methods in the *social* sciences at least by denying that the discovery of generalisations for explaining and predicting phenomena is an appropriate aim for the social scientist. In support of this they have referred, variously, to the subjective meaningfulness of social action, to the 'value-ladenness' of social scientific judgments, or simply to the sheer complexity of social phenomena.

I shall not follow this line of argument. As it happens, I do not think that any of the considerations mentioned provides any conclusive reason for supposing a qualitative distinction between the social and natural sciences. But instead of arguing this I shall simply *assume* that the social and natural sciences alike are ultimately aiming at the discovery of factual generalisations, and will try to show that even *within* this assumption a defence can be found for the use of ideal type methods. Indeed, the essence of my arguments will be that the use of ideal type methods is merely a special case of what must be generally acknowledged to be an important aspect of practice in the natural sciences. But first I must clear away some confusions.

2 'The adequacy of a hypothesis does not depend on the realism of its assumptions, but merely on the accuracy of its empirical consequences.' This dictum is strenuously defended by Milton Friedman in his [1953] and is often invoked in defence of ideal type methods (in particular in defence of the use of abstract models in economics). However, there is a basic difficulty involved in knowing what is meant by 'realism of assumptions' here. This phrase seems to have been intended, and understood, in (at least) two different ways. When these are distinguished it becomes clear that Friedman's dictum offers no help to the defender of ideal types.

In one sense, ideal type generalisations have 'unreal assumptions' simply insofar as their antecedent terms fail (by definition) to apply to any actual situations. However, generalisations characterised by such 'unreal assumptions' can scarcely be defended by reference to the 'accuracy' of the empirical predictions they yield, for, as pointed out in the previous section, such generalisations will never yield any empirical predictions, accurate or otherwise.

On the other hand it is possible to understand 'unreal assumptions' as referring, not to the antecedent terms of ideal type generalisations, but to such generalisations *in toto*, and thus to read Friedman's dictum as suggesting that the acceptance of ideal type generalisations might, in spite of their *falsity*, be justified in terms of their deductively implying other generalisations which are themselves capable of direct empirical corroboration. However, it is hard to see how this can help. For what is wrong with ideal type generalisations is *not* that they are false, but that they are *unfalsifiable*. The admitted fact that any attempt to apply their antecedent terms to specific situations will result in false singular statements does not show that ideal type generalisations themselves are false; rather it shows that there is no possibility of their truth or falsity being empirically indicated. Once we are clear that ideal type generalisations are unfalsifiable, rather than false, we can see that there is no real possibility of their being justified by their deductively implying other generalisations whose empirical accuracy might be corroborated. The direct consequences of a generalisation whose conditions of application are never instantiated will in general themselves be generalisations with non-instantiated antecedent terms.

It might be argued nevertheless that ideal type hypotheses do play a part in the derivation of empirically testable generalisations when they are conjoined with other assumptions. Thus consider the following kind of argument:

(1) Ideal gases (those composed of perfectly elastic molecules with zero mass and volume) satisfy Boyle's law.

(2) If the temperature is above and the pressure below certain (specified) limits, hydrogen is an ideal gas.

Therefore

(3) Within those limits of temperature and pressure, hydrogen satisfies Boyle's law.

Here (1) plays a part in deriving (3), in that (3) follows from (1) and (2), but not from (2) alone. However, we cannot simply therefore take the empirical acceptability of (3) to justify the acceptance of the ideal type hypothesis (1). For the assumption (2) is manifestly false. We do not in general consider statements to be responsible for what they imply when conjoined with admitted falsehoods, and consequently do not take the truth of such implications to support them. (There would be no limit on what could be justified if we adopted such a principle.) Clearly something more needs to be said about the status of (2) in the above argument if the acceptability of the argument's conclusion is to count as lending support to (1). One thing that might be done (something like this is suggested by Friedman) is to replace (2) by

(2)′ Within the specified limits of temperature and pressure hydrogen behaves *as if* it were an ideal gas.

Since something 'behaving *as if*' it were an ideal gas presumably requires only that it does what ideal gases are supposed to do (and not its actually being one), there is no longer any bar to (2) being credited as a seriously accepted assumption. However, this ploy of introducing 'as if' premises simply makes the whole line of defence under consideration irrelevant to the justification of ideal type methods. For note that if the sample argument is to go through with (2)′ in place of (2), then (1) must be replaced by

(1)′ Anything which behaves *as if* it were an ideal gas satisfies Boyle's law.

But (1)′ cannot be held to be an ideal type generalisation at all, since there *are* things which behave *as if* they were ideal gases (at least insofar as (2)′ is acceptable). And in any case (1)′ is analytic, and so does not any more play any real part in the derivation of (3), which is now directly derivable from (2)′. Such objections as these can clearly be made to any attempt to justify ideal type methods by reference to 'as if' interpretations.

Before proceeding further it is worth noting that the unfalsifiability that is characteristic of genuine ideal type generalisations need not be due to their being accepted as analytically true. That is, it is not generally the case with ideal type generalisations that the consequent term's being satisfied by some entity is a logically necessary condition for the antecedent term's applying to that entity. For instance, somebody who asserts that 'In pure vacuums bodies fall with constant acceleration' need not have as part of his ultimate criteria for something being a pure vacuum that bodies fall in it with constant acceleration. Yet his assertion will still be untestable, simply for lack of any pure vacuums. The unfalsifiability of ideal type generalisations is not so much an unfalsifiability in principle as an unfalsifiability in practice. (It might be objected here that the acceptance of an ideal type generalisation *does* generally provide a new criterion for applying the terms involved, and thus that ideal type generalisations are indeed unfalsifiable in principle. This objection raises the important question of whether it is possible to distinguish between analytic and synthetic lawlike generalisations. However, this objection and the issues it raises are in no way peculiar to ideal types and so I shall avoid discussing them further here.)

It is perhaps also worth noting that ideal type generalisations are not the same thing as '*ceteris paribus* laws'. The latter are typically stated with non-ideal concepts as antecedent terms, but are rendered unfalsifiable by the addition of a '*ceteris paribus*' clause which can be invoked whenever a negative instance appears. Ideal type generalisations, on the other hand, being unfalsifiable to start with, do not require any such clause to make them so. ('*Ceteris paribus*' laws and ideal type methods do raise similar problems; indeed it seems to me that the usefulness of generalisations with ideal type concepts as *consequent* terms depends precisely on the availability of '*ceteris paribus*' clauses. Unfortunately space restricts me from discussing '*ceteris paribus*' laws any further here.)

3 It might be thought that the utility of ideal types can be accounted for easily enough simply by understanding ideal type concepts as implicitly designating not only those (non-existent) situations that satisfy them exactly, but also as applying to any situations that 'approximate' to the ideal type.

However, this account can be faced with an awkward dilemma. Is it in general specified what degree of approximation is required for something to count as an instance of (what is intended by) the ideal type, or is it not? If it is *not*, then it can be argued that ideal type generalisations are quite vacuous—for in the absence of such specifications it will be quite arbitrary what are to be taken as instances (whether negative or positive) of such generalisations. If, on the other hand, such specification *is* laid down, it is difficult to see why any recourse to ideal type methods is needed at all: why, in such cases, is the generalisation not simply asserted in the way it is meant to be understood? We would then have a generalisation framed in terms of a normal descriptive concept more general in scope than the original ideal type, and all difficulties would disappear. But to take this horn of the dilemma is scarcely a solution, for it simply denies to ideal types just that characteristic which makes them problematic in the first place.

These considerations also show what is wrong with the view that the usefulness of ideal type generalisations lies in their ability to yield predictions and explanations when 'simplifying' assumptions about particular situations are made. Suppose someone 'assumes' that a given medium is a perfect vacuum, and thence predicts that a body falling through it will have constant acceleration. The obvious objection to such a procedure is that his prediction is based on a premise known to be false. Since we do not in general credit such predictions, there must be some constraints on when this procedure can be used. Presumably the justification in a particular case would be that the simplification involved deviates so little from reality that we can rely on the resulting prediction to be approximately true. If this defence is to carry any weight, however, there will have to be definite limits on the extent of simplification allowed and the degree of approximation to be expected in the prediction; and it will have to be accepted that the ideal type generalisation 'holds good' within those limits. But this is just to say that the generalisation involved is not essentially ideal typical after all, but, as above, implicitly an ordinary generalisation framed in non-ideal terms.

4 The most sophisticated defence of ideal type methods is based on a comparison of ideal type concepts and theoretical concepts. According to this view, which can be found in Ernest Nagel's [1963] (and, less centrally, in Carl Hempel's [1965]) the acceptance of an ideal type generalisation is justified only when the generalisation is the limiting case of a more general system of hypotheses, which are capable of, and have received, independent empirical corroboration. For instance, it is argued that Galileo's law of free fall in a perfect vacuum is acceptable because it is the limiting case of a system of generalisations which describes how bodies do fall in actually occurring situations. It is the limiting case in the sense that the system of generalisations in question can be summarised by a simple continuous function which gives the acceleration of a falling body in terms of the density and elasticity of the medium and the shape and speed of the body, and which specifies that the acceleration tends to a constant as the density of the medium tends to zero.

What underlies Nagel's claim that ideal type generalisations are justified in such cases is the notion that terms signifying ideal type concepts are a particular kind of theoretical term. By 'theoretical' term is meant an expression the applicability of which cannot be decided by direct observation, but only by means of inferences based on 'mixed' generalisations linking the theoretical terms to observational terms. In general the utility of generalisations involving theoretical terms is argued to be that they allow a large number of observational generalisations to be summarised in a simple and systematic manner: relatively few theoretical postulates and 'mixed' generalisations will generate as deductive implications a wide range of observational hypotheses. Thus the essence of Nagel's argument is that ideal types, like theoretical terms in general, can play an essential role in frameworks of postulates which allow an economical and systematic statement of what would otherwise be an unmanageable mass of observational generalisations.

However, the assimilation of ideal types to theoretical terms involved here can be shown to be illegitimate. When this is done it will be seen that the Nagel account of ideal types is unsound. What is characteristic of ideal types is not that they can only be applied indirectly—rather it is that they can't be applied at all, for lack of any instances. 'Ideality' and 'theoreticity' are quite distinct: a theoretical term certainly does not need to lack instances ('radioactive', 'monarchy'); and conversely we can conceive of ideal types which are observational, in that if there were any instances it would be possible to recognise them as such by direct sensory inspection ('black tulip'). The clear distinction between ideal types and theoretical terms should not be obscured by the fact that many of the terms used in science are both ideal typical and theoretical.

Once we are clear about this distinction it is possible to see that the Nagel account of ideal types is unsound. What is held to justify the acceptance of theoretical postulates is that, together with mixed generalisations, they can provide economical sets of premises from which many observational generalisations can be deduced as consequences. But it is essential to theoretical postulates making a real contribution to such deductions that the concepts they involve be non-ideal; to put it crudely, they must be about theoretical aspects of actual situations. For, as argued above in section 2, ideal type generalisations, being nowhere applicable, cannot play any real part in the deduction of empirically testable consequences. True, as Hempel and Nagel point out, an ideal type generalisation can be a limiting case of a general function stating the relationship between specified deviations from the ideal type and the phenomena which will occur in such deviating situations. But the ideal type generalisation will not add anything to the empirical content of a statement of such a functional relationship. This is made clear by the fact that the empirical implications would remain the same if we replaced the function in question by another differing only in that it was undefined, or had some singularity, at the limiting (ideal) case. Although it would be unnatural (and probably counter to any principles of inductive logic there might be) to leave the function undefined for the ideal case, or to postulate a singularity there, the possibility does show that the acceptance of ideal type generalisations cannot satisfactorily be justified by supposing them to be *necessary* for economically formulating a system of postulates with observational consequences.

This point can be elaborated by comparing the use of a system of theoretical postulates to make an explanation or a prediction with the analogous use of a function with an ideal type generalisation at the limit. In the former case we reason from established initial conditions, via mixed and theoretical generalisations, to the state of affairs to be predicted or explained. The theoretical generalisations are essential to the validity of this deduction in that without them the deduction would not go through. When in the latter case, on the other hand, we infer from the extent of the deviation of the initial conditions from the ideal type case, what then will (or had to) happen, the inference is sustained solely by what the function says about *that* kind of deviation, independently of whatever it might assert for the ideal type case. Any actual prediction or explanations could here be made equally well even if we dispensed with the ideal type generalisation itself.

It might be doubted whether the dispensability of ideal type generalisations for predicting and explaining is in fact a sufficient reason for discounting their role in the formulation of empirically contentful systems. For systems of theoretical postulates are also in principle dispensable, notwithstanding their practical usefulness in making inferences from one observed fact to another. Craig's theorem shows that for any system of postulates involving theoretical terms there is another with the same empirical content involving observational terms only. However, in the present context this is not to the point. Scientists use theoretical postulates, in spite of their dispensability in principle, because their replacement by empirically equivalent systems of observation generalisations would involve great loss of convenience and simplicity—the observational replacement invoked by Craig's theorem cannot in general be axiomatised by a finite schema. Thus practising scientists have good reason to rely on theoretical postulates in predicting and explaining. This is not so with ideal type generalisations. What I have argued is that ideal type generalisations do not *even in practice* add to the explanatory or predictive power of systems of generalisations. No additional complexity would result from their abandonment, and so there is no question of their practical utility justifying their acceptance.

5 However, it is possible to derive a justification of ideal type methods from a comparison of ideal type and theoretical concepts, if we focus on a different aspect of theoretical procedures from the one considered so far. Facilitating the economical summarising of observable generalisations is not the only role that theoretical methods play in scientific practice. They are also significant in affecting the development of science over time. For often a scientist accepting certain postulates at a theoretical level will have a commitment to these postulates in themselves, over and above any commitment he may have to the empirical generalisations they are taken to imply, a commitment which will be manifested in his seeking to substantiate precisely those observational hypotheses which would lend

support to his favoured theoretical assumptions, and even, if necessary, in his revising 'mixed' generalisations in such a way as to ensure that the observational data do fit in with those assumptions. In general this kind of procedure lends direction to scientific research; moreover it can be further justified in that it provides some guarantee that the observational generalisations that are admitted to the corpus of scientific knowledge will actually be susceptible of economical incorporation into theoretical systems. Imré Lakatos, in his [1970] has given a general characterisation of this aspect of scientific practice. Lakatos argues that the history of science displays a succession of competing 'research programmes', each consisting of a framework of fundamental principles which over time will find expression in a developing sequence of systems of observable regularities. More precisely, a 'research programme' is characterised by a 'negative heuristic' or 'hard core', a set of basic postulates which are, by fiat, accepted as unfalsifiable, together with a 'positive heuristic', which is 'a partially articulated set of suggestions or hints on how to change, develop the "refutable variants" of the research programme' (p. 135).

It seems to me that the most fruitful way of understanding ideal type methods is as another application of the methodology Lakatos describes and advocates. An empirical system might be simple and economic either because it is derivable from a small number of theoretical and mixed generalisations, or because it can be summarised in a simple functional relationship. An ideal type approach to some field amounts to a 'research programme' directing the scientist towards a system with the latter kind of economy. The adoption of an ideal type generalisation can be seen as a special case of the adoption of a 'research programme' as follows: the ideal type generalisation itself is the 'negative heuristic', a basic principle which is itself unfalsifiable, but which together with the 'positive heuristic' generates a series of empirical generalisations. The 'positive heuristic' would be some 'partially articulated suggestions' of the following kind: suggestions about the limits within which situations approximating to the ideal type situation will approximately satisfy the consequent term of the ideal type generalisation; suggestions about the dimensions along which such approximations can fruitfully be differentiated; suggestions about the kinds of generalisations that might relate specific approximations thus differentiated; *etc.* Somebody espousing an ideal type generalisation would be committed to the existence of some simple but unspecified function having as values deviations from the consequent term and amongst its arguments deviations from the factors mentioned in the antecedent term, with the property that the consequent term would be the limit of this function as the deviations from the factors in the antecedent term tended to zero.

The distinction between ideal types and theoretical terms emphasised above does not mean that ideal types cannot play a similar role to theoretical terms in research programmes. For what is at issue here is not the ability of ideal type generalisations to contribute to the empirical content of systems of generalisations considered synchronically, but whether they can play a part in the diachronic development of such systems. Even if, as I have argued, the acceptance of ideal type generalisations can add nothing to the empirical content of systematic frameworks of generalisations, it is still possible that they can be essential in the production of such frameworks. If this is correct, then ideal types should be seen as providing the kind of ladder which may as well be thrown away once it has got one where one wants to go.

The account of ideal types I am advocating differs from that offered by those who assimilate ideal types to theoretical terms particularly in respect of one important consequence. On their view, the acceptance of an ideal type generalisation is unjustified unless it is known to fit into an independently corroborated system as a special case. As Hempel in particular is concerned to point out, this would deny any sound basis to nearly all the ideal typical assertions put forward by social scientists—for in general such assertions are patently not set within the context of a framework of empirically confirmed generalisations. The view adopted here does not require any such methodological intolerance. For it suggests that the acceptance of an ideal type generalisation is not to be justified by showing that it fills an otiose position in some framework of generalisations, but, if at all, by the possibility that it might lead to the elaboration of such a framework.

Of course the argument I have put forward does nothing to show that social scientists will meet anything like the success of their natural counterparts in finding such frameworks of generalisations. But it does imply that, if that is their aim, the adoption of ideal type methods is a sensible way to go about it and nothing to be ashamed of.

University of Reading

REFERENCES

FRIEDMAN, M. [1953]: 'The Methodology of Positive Economics', in *Essays in Positive Economics*.

HEMPEL, C. [1965]: 'Typological Methods in the Natural and Social Sciences', in *Aspects of Scientific Explanation*.

LAKATOS, I. [1970]: 'Falsification and the Methodology of Scientific Research Programmes', in I. Lakatos and A. Musgrave (*eds.*): *Criticism and the Growth of Knowledge*.

NAGEL, E. [1963]: 'Assumptions in Economic Theory', *American Economic Review, Supplementary Volume*, **53**, 1, pp. 211–19.

Discussions

A NOTE ON VERISIMILITUDE

1 *Criticism of Tichý's Proposal.*
2 *Distance from the Truth.*
3 *Approximation to Truth by Numerical Theories.*
4 *On Problems Raised by Miller.*
5 *Concluding Remarks.*

The four recent discussion notes on verisimilitude in this *Journal* by Pavel Tichý ([1974]), John H. Harris ([1974]), and David Miller ([1974*a*], [1974*b*]) have had a somewhat shattering effect on the situation of the theory. All of these papers are in my opinion very important.

The present note which is greatly indebted to David Miller for suggestions, improvements, and for cuts, contains, apart from some critical remarks on the preceding notes, a few positive hints (developed mainly from a suggestion by Stefan Mazurkiewicz [1934]).

Tichý proposes in his section 4 a new definition of verisimilitude applicable to very simple languages and he announces an extension of this to richer languages.[1] I am extremely anxious to have the status of verisimilitude established, and for this reason I very much hoped that Tichý's theories would be acceptable. Miller ([1974*a*], pp. 175–7) has raised very interesting objections to Tichý's theory: he showed that it made verisimilitude dependent on language in a manner that would clearly be unacceptable in a final theory. I hoped however that once we had obtained a definition satisfactory with respect to one language, it would be possible to obtain a measure of verisimilitude which like truth would be invariant with respect to translations into other languages. I still think that this might be a promising way. But I think now that Tichý's proposed metrical definition suffers from further inadequacies (inadequacies which, incidentally, are remarkably similar to those which he himself has pointed out in my own metrical definition).

In what follows I shall first point out these inadequacies in Tichý's definition. Secondly I shall refer to an alternative qualitative theory of distance from the truth, proposed by Miller in an as yet unpublished paper entitled 'On Distance from the Truth as a True Distance'. I shall show that my old proposals can be recovered from a variant of this theory, due to Mazurkiewicz; and I shall then sketch what happens when the theory is put in quantitative form. Finally, I shall discuss a third proposal which seems to me as satisfactory as any.

1 *Criticism of Tichý's Proposal*

After pressing home very effectively his criticism of my proposals, Tichý proposed in section 4 of his [1974] a definition of the distance of an arbitrary sentence *a* from the truth, as follows. The truth is represented by one of the strongest true statements of the language considered; in the case of his language L which contains only three atomic sentences, h, r, w (it is hot, it is raining, it is windy) it is assumed that $h.r.w$ is true (and therefore the strongest true statement from which all other true statements such as $h \vee r'$ follow, where r' is the negation of r). In order to determine $d_T(a)$, the distance of a from the truth, Tichý proposes that we put a into disjunctive normal form, count the negation signs, and divide by the number of the (conjunctive) constituents of the disjunction (*i.e.* the number of the disjunction signs plus 1).

---

[1] I gather that Tichý's new paper ([1976]) is to appear in the March issue of this *Journal*.

Thus we get for a language like L with 3 atomic sentences

$$d_T(t) = d_T(h.r.w) = 0 < d_T(x) < 3 = d_T(h'.r'.w').$$

The true statements are those in whose disjunctive normal form the statement $t = h.r.w$ appears as one of the constituents; for example, the true statement h will be written in disjunctive normal form:

$$h = h.r.w \vee h.r.w' \vee h.r'.w \vee h.r'.w'.$$

Thus it contains 4 negation signs and 4 constituents; and $d_T(h) = 1$. (Of course, $d_T(h) = d_T(r) = d_T(w)$.)

The tautology of L consists in disjunctive normal form of all 8 constituents. There occur 12 negation symbols in it. Accordingly, we have

$$d_T(taut) = 12/8 = 1.5.$$

Tichý, it seemed to me, implied that he accepted my view that a *true* theory a is *the nearer to the truth the stronger it is*; thus its distance $d_T(a)$ becomes 0 when a becomes logically equivalent to the strongest theory, $h.r.w$. Since he did not seem to say so explicitly, I wrote to him (on 16 July 1975), raising the point of comparing the distance from the truth of *true* theories (though agreeing that the distance from the truth of false theories is more important). In his reply (dated 30 July 1975) Tichý wrote, as I expected, that a true sentence is simply the closer to the truth the stronger (or logically less probable) it is—exactly as in my old theory (and as in my present theory). Professor Tichý also sent me the actual figures for the value of d_T of eleven true sentences.

I accepted all this, and hoped that his theory would work. However, quite recently I found a number of trivial counterexamples.

Let us start from the statement in normal form

$$a = h.r.w \vee h'.r'.w',$$

and from the sequence of statements each of which is logically derivable from (and therefore weaker than) its predecessor:

$$b = a \vee h.r.w'$$
$$c = b \vee h.r'.w$$
$$d = c \vee h'.r.w.$$

We obtain for these four statements the Tichý-distances from the truth:

$$d_T(a) = 1.5 \qquad d_T(b) = 1.3 \qquad d_T(c) = 1.25 \qquad d_T(d) = 1.20.$$

Thus the distances from the truth of these true statements *decline* with their declining logical strength, although they ought to *increase* with declining logical strength. This surely happens against Tichý's own intuition and intention; and it shows that Tichý's definition is inadequate.

There are many other similar anomalies; for example:

$$1 = d_T(h.r.w') = d_T(h.r.w' \vee h.r'.w) = d_T(h.r.w' \vee h.r'.w \vee h'.r.w).$$

All these theories are false. They have all equal d_T, even though they show decreasing strength.

Moreover, the tautology ($d_T(taut) = 1.5$) which is of all true theories 'the lousiest' (to use Tichý's term) has the same distance from the truth as the relatively good and strong theory $h.r.w \vee h'.r'.w'$, or the theory $h \vee r'$. It is actually closer to the truth than is the true non-trivial theory $h.r.w \vee h'.r'$.

All this shows that Tichý must be mistaken when he claims without proof ([1974], p. 159) of his definition: 'It is easily seen that such a definition meets all intuitive requirements.'[1]

2 *Distance from the Truth*

By the time I found this trivial (but I suspect irreparable) inadequacy in Tichý's proposal, I had started on a plan of making Tichý's or a similar definition invariant with respect to a translation into another language, such as Miller's ([1974*a*], pp. 175–7) *h*—*m*—*a* language. This now became pointless. But another and more general approach seemed more promising. This was based on a paper by Stefan Mazurkiewicz [1934], a paper I had known well for many years, and to which I had referred before (in my [1972], p. 333; see also my [1938], *i.e.* my [1959], p. 320.)

Mazurkiewicz's paper contains a simple definition and some theorems about what he called the '*stochastic distance*' (or probabilistic distance) between two theories or deductive systems. (Mazurkiewicz's probabilistic theory is based on Tarski's calculus of deductive systems.) I remembered this on the occasion of a conversation with David Miller who told me that he was working on a paper about distance from the truth, and that his distance satisfied the triangular inequality. When I heard this I said: 'Mazurkiewicz defines a distance between theories which satisfies the triangular inequality: this must be the solution to our problem.'

Whether this is indeed 'the' solution to our problem I doubt very much. However, in developing Mazurkiewicz's definition I was led quite unexpectedly to what are essentially my old proposals. These proposals are of course open to some of the old objections of Miller, of Tichý, and of Harris. But it seems to me that these objections look a little different now.

Mazurkiewicz introduces in his paper [1934] an axiom system for probability, based on the calculus of deductive systems of Tarski [1935], [1936*a*]. He defines the stochastic or probabilistic distance between two deductive systems a and b by

2.0 $$d(a, b) = p(a.b' \vee a'.b) = p(a.b') + p(a'.b).$$

Since $a.b' \vee a'.b$ is also called, in Boolean algebra, 'the complete disjunction of a and b' or 'the symmetric difference of a and b', we can say that Mazurkiewicz defines the probabilistic distance of a and b as the probability of their complete disjunction or symmetric difference. Mazurkiewicz gives an acknowledgement to Fréchet [1930], [1931]. David Miller has drawn my attention to a similar definition in Nikodym [1930]. A Venn diagram shows the intuitive character of the complete disjunction as a definition of qualitative distance.

---

[1] I sent an earlier draft of the present note to Pavel Tichý, who replied that he upholds his definition in the face of my examples (which he acknowledged as correct) and that his earlier letter to me had been hasty: he had been wrong in agreeing that the verisimilitude of true statements increases with content. But what then, I ask, about his claim, here quoted, concerning 'all intuitive requirements'? And what about the reliability of his intuitions concerning such a clear-cut point?

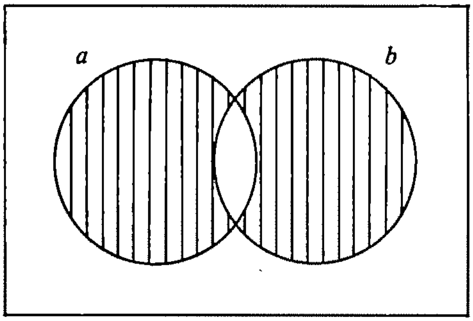


Fig. 1.

Mazurkiewicz shows that $d(a, b)$ satisfies the following conditions 2.1 to 2.9 of which 2.1 to 2.4 may be taken (following Fréchet) as characteristic of any distance function

2.1 $$d(a, b) = d(b, a)$$
2.2 $$d(a, a) = 0$$
2.3 $$d(a, b) + d(b, c) \geqslant d(a, c).$$

This is the triangular inequality. It can be easily proved algebraically or with the help of a Venn diagram (Fig. 2).

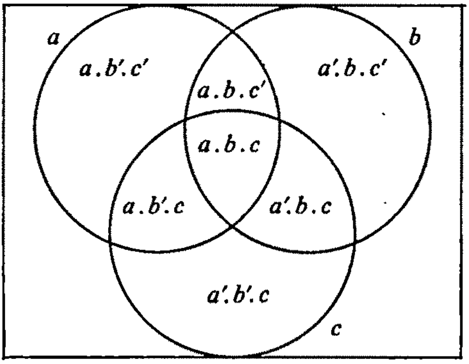


Fig. 2.

We have, of course, by 2.2,

2.4 If $a = b$ then $d(a, b) = 0$
2.5 $d(a, b) \geqslant 0$ (by 2.1 to 2.3)

(Incidentally, Mazurkiewicz explicitly does *not* assume the validity of

If $d(a, b) = 0$ then $a = b$;

he rather introduces a new relation, the congruence of a and b, in symbols $a \equiv b$, and writes: If $d(a, b) = 0$ then $a \equiv b$.)

The following formulae are also characteristic of Mazurkiewicz's probabilistic distance function:

2.6. $d(a, b) = d(a', b')$

2.7 $d(a', b) = d(a, b') = 1 - d(a, b)$

2.8 $d(a, a') = 1$

2.9 $d(a, b) < 1$.

This theory of distance between elements of a Boolean algebra can be transscribed into a qualitative form, provided we give 0, 1, +, —, < and > Boolean rather than arithmetic interpretations. (See Miller's 'On Distance from the Truth as a True Distance' for references to the literature about this.) Mazurkiewicz himself does not introduce a qualitative (or quasi-metrical) distance; nor does he mention the distinction between a qualitative distance definition and the metrical or probabilistic definition which he introduces. Yet it is a fact that part I of his paper is confined to the finitist theory of probability (to my knowledge part II was never published), so that his algebra of systems is Boolean; and it can easily be shown that the Boolean operation of symmetric difference, $D(a, b) = a \,.\, b' \vee a' \,.\, b$, can be interpreted as a *qualitative distance function* which satisfies all the metrical or numerical formulae 2.1 to 2.9, satisfied by the function $d(a, b)$, if we interpret, where appropriate, > and <, 0 and 1, and + and — in the Boolean sense.

The qualitative formulae (with D instead of d) may be called 2.10 to 2.19. They can all be proved very simply and intuitively with the help of Venn diagrams. The probabilistic formulae 2.1 to 2.9 would follow then immediately from 2.11 to 2.19. I shall only write down 2.10 and 2.11.

2.10 $D(a, b) = a.b' \vee a'.b$

2.11 $D(a, b) = D(b, a)$.

2.12 to 2.19 will be omitted to save space.

In what follows I too will confine myself to the Boolean case rather than apply Tarski's calculus of systems (of which I have made use before [1972]).

We introduce the following definitions of the *qualitative and quantitative distances from the truth.* (T is the set of true statements of L; t is the strongest true theory. We assume here that t exists; that is, that T is axiomatisable.)

2.20 $D_T(a) = D(a, t) = a.t' \vee a'.t$

2.21 $d_T(a) = d(a, t) = p(a.t') + p(a'.t)$.

In view of the fact that we have the formulae 2.6 and 2.16 we can introduce the following definitions for *qualitative and quantitative nearness functions* (truthlikeness, verisimilitude):

2.22 $N_T(a) = 1 - D_T(a) = D_T(a')$

2.23 $n_T(a) = 1 - d_T(a) = d_T(a')$.

If we look carefully at the definitions above of the qualitative distance and nearness functions we can see that the essential point is the Boolean character of the space of *values* of D and N. It is not necessary that this Boolean algebra should actually be the same Boolean algebra as the one that is composed of the arguments (statements). In his paper Miller considers two possibilities only: (*i*) the algebra of values is simply the algebra of statements; (*ii*) the algebra of values is the algebra of ranges (or models, or interpretations). (In the case of sufficiently simple languages these two algebras are isomorphic.) It is of some interest, however, to consider another Boolean algebra than the algebra of values, namely the ***algebra of sets of statements***. In this algebra the operation of complementation has a markedly different force from its force in the algebra of statements. For example, in the latter algebra t' is the negation of the strongest true statement t; it is therefore the weakest of all false theories. (It is what Tarski called an ***irreducible*** theory.) But if we identify t with the set of true statements T, its complement (relative to the set of all statements) is the set F of false statements. The qualitative ***nearness to the truth*** (verisimilitude) of a in this algebra is represented by the shaded parts of Fig. 3.

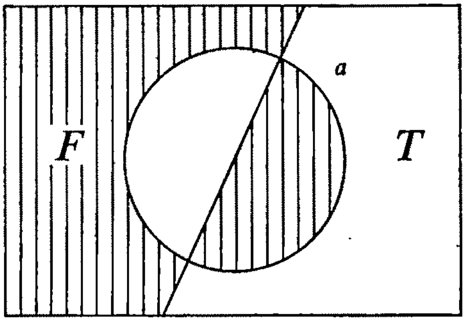


Fig. 3.

Now these two shaded areas, when read from right to left, represent ***the truth content of the theory a***, which I have previously written $Ct_T(a)$, and what remains in the set F of false statements when ***the falsity content*** $Ct_F(a)$ ***of the theory a*** is removed. Thus we can write

2.24 $$Vs(a) = Ct_T(a)+F-Ct_F(a)$$

for ***the verisimilitude of the theory a***. Since the set F is constant, it follows at once that

2.25 $$Vs(a) < Vs(b) \text{ iff } Ct_T(a) < Ct_T(b) \text{ and } Ct_F(b) \leq Ct_F(a),$$
$$\text{or: iff } Ct_T(a) \leq Ct_T(b) \text{ and } Ct_F(b) < Ct_F(a).$$

This is the old proposal of my [1963], p. 233. It is clear that Miller's and Tichý's main criticisms of that proposal must apply with unmitigated strength to the new one. But since the new proposal is based on a general definition of distance between theories, and thus, as it were, gives a theoretical explanation of the old one, I do not feel that it is quite useless to mention this derivation.

I return now to the distance function that Miller considers. He shows in his paper that the theory of distance that it generates is a somewhat rudimentary one, the main consequences being the following three.

First: if a and b are both true statements and $a < b$ then $D_T(a) < D_T(b)$.
Second: if a and b are both false statements and $a < b$ then $D_T(a) < D_T(b)$.
Third: if a is a false statement then $D_T(a \vee t) < D_T(a)$.

Of these the first is what one would expect of comparable true theories (and it is the result that Tichý's theory fails to deliver). The third consequence is very satisfactory since it shows that in some cases at least the weaker of two comparable theories is the one nearer to the truth. The truth content of a false theory is undoubtedly weaker than the theory itself, but it is also, intuitively speaking, closer to the truth, since it contains all the true consequences that theory contains, and none of the false ones. (Tichý's theory also gives this result.) But the second consequence may give us pause. To it Tichý's comment ([1974], p. 157, n. 2) applies: 'If this definition were adequate it would be child's play to increase the verisimilitude of any false theory A: it would suffice to add to A an arbitrary sentence which does not follow from it.'

David Miller tells me that he finds this criticism unconvincing, despite its initial plausibility. But what I wish to do now is to show how elegantly all three of the consequences listed above can be presented if we move back from qualitative distance functions to *quantitative distance and nearness functions*. In particular I wish to return to the function n_T defined in 2.23 *above*.

Our results can be summed up in two theorems. Because of the role played in what follows by $ct(t') = p(t)$, I will introduce for this constant (it is a constant depending on the language under consideration) the symbol q, and for $1-q$ the symbol q':

2.26 $$q = ct(t') \qquad q' = p(t')$$

Since t' is the weakest false statement of the language it is one of those statements called by Tarski 'irreducible'. We can assume that our language (whether finite or not) contains irreducible statements, and we can attribute to each of them the content measure q.

THEOREM 1. If a is true and b is false then

$$n_T(a) = ct(a)+q;\ n_T(b) = ct(b)-q.$$

Thus if a and b have equal content

$$n_T(a) = n_T(b)+2q.$$

Proof: If a is true, it follows from t, and $p(a.t) = p(t)$, and $p(a'.t') = p(a')$, since t' is the weakest false theory. If b is false, $p(b.t) = 0$ and $p(b'.t') = p(b')-p(b'.t) = p(b')-p(t)$, since $p(b'.t) = p(t)$ if b' is true.

The theorem follows immediately.

THEOREM 2. If a is true and b is false and they have equal content (but of course different n_T), then $a\,.\,t'$ is false and $b \vee t$ true and these two have equal n_T (but different content); more precisely, we get, if a is true and b is false,

$$n_T(a.t') = n_T(b \vee t) = n_T(a)-q = n_T(b)+q = ct(a) = ct(b);$$
$$ct(a.t') = ct(b \vee t)+2q$$

Theorems 1 and 2 are illustrated in the following diagram.

As the diagram indicates, $n_T(t) = 1$, $n_T(t.t') = q'$, $n_T(t \vee t') = q$, $n_T(t') = 0$. All true theories (like a) are to be found on a straight line $n_T(x) = ct(x)+q$; all false theories (like b) lie on a parallel lower line $n_T(x) = ct(x)-q$. The theory $a.t'$ is, of course, the weakest false theory from which a follows, and $b \vee t$ is the strongest true theory which follows from b.

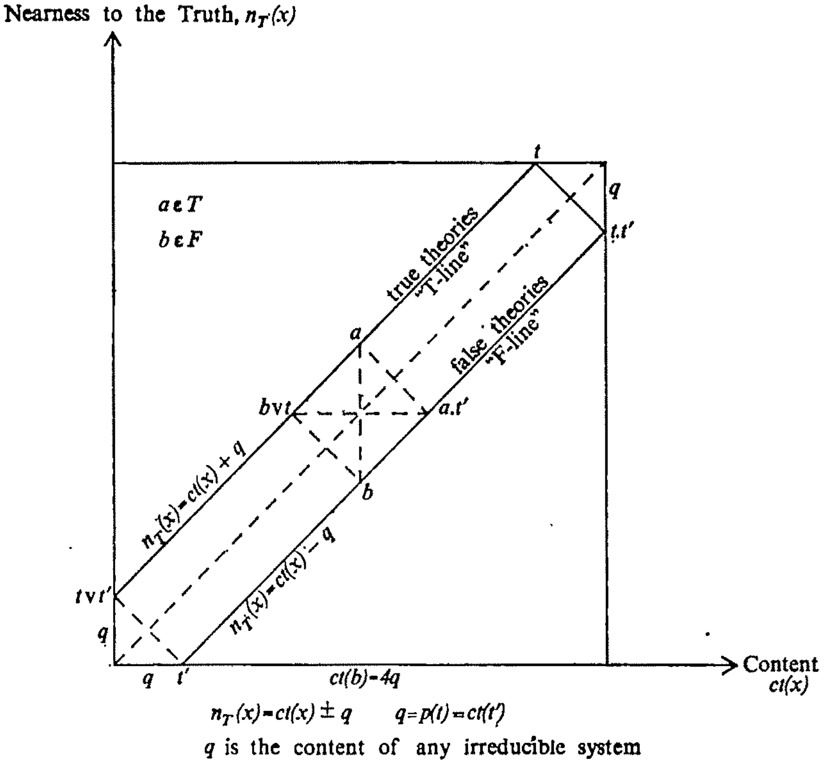


Fig. 4. The diagram does not distinguish between any two theories of equal truth value and equal content.

We can now see precisely the advantages and disadvantages of our proposed definition of nearness to truth.

(1) For true theories it works perfectly well; here increase of content is clearly increase of truthlikeness.

(2) For the comparison of false theories it does not, at first sight, work so well; for we should like the conjunction *a.b* of a true theory *a* (especially if *a* is strong) and a false theory *b* (especially if *b* is weak) to be nearer to the truth than an equally strong conjunction *a.b* of two false theories. It seems, however, that this intuitively appealing demand (in which I originally believed) *cannot* be met. For the resulting theory *a.b* is in both cases false, and it can in both cases be turned into a true theory simply by disjunctively adding *t*, so that we get (*a.b*) ∨ *t*. In other words, whether *a* is true or false makes no difference (1) to the distance to the nearest and strongest entailed true theory (the truth content of *a.b*), or (2) to the logical strength of this true theory. I now think that my intuition misled me, and that it must be corrected in the light of these results.

(3) The proposal allows of some good applications in connection with the *comparison of true theories with false ones*. Not only are equally strong true and false theories always distinguished with respect to truthlikeness in the way expected, but we have:

THEOREM 3. Although a conjunction $a.b$ will (because of its increased content) in general be nearer to the truth than either of its components, this is not always the case if $a.b$ is false and a is true. For in this case it may happen that

$$ct(a.b) > ct(a) \;\&\; n_T(a) > n_T(a.b)$$

(Of course, there are also cases in which the second sign '$>$' is replaced by an equality sign.)

Thus it is sometimes possible to compare the verisimilitude of a false theory with that of other theories, either true or false.

3 *Approximation to Truth by Numerical Theories*

This, I think, is about all we can get from the proposal to define nearness to truth as nearness to the set T of all true statements. As a first step beyond this, I suggest that this set T is too big. We may admit into our universe of statements only such statements as we conjecture to be relevant; relevant, that is, to the problem situation in hand. (See my [1963], n. 24 on p. 241.) This contraction of the universe of admissible statements *may* be interpreted as the introduction of a language which is confined to the problem in hand (such as Tichý's closed language *h—r—w* which is confined to the three problems 'Is it hot?'—'Is it raining?'—'Is it windy?'). However, I prefer not to think in terms of a closed artificial language but in terms of a language into which we can freely introduce new problems and new statements whenever we conjecture that this is relevant to our problem situation. At the same time, I wish to take a leaf out of Dr Harris's book and allow for the introduction, in the place of T (the complete deductive system of all true statements), of *weaker comparison theories* which are (true) answers to the problems under consideration. I will denote (following Harris) the set of the comparison theories under consideration by the capital letter C. I will assume that C is true, and I will ask for $n_C(a)$ in place of $n_T(a)$.[1]

By confining ourselves to *relevant* conjectures we can, I think, solve the problem of how to avoid strengthening a theory by inserting just any stray irrelevant conjunct. Any such procedure would have first to be justified as relevant to a solution of one of the problems in hand.

One significant point should be mentioned here. Our diagram with the two lines parallel to the diagonal representing true and false theories respectively applies fully to an arbitrary comparison theory C, provided the parallel lines are now taken to represent C-true and C-false theories only. But if a C weaker than T is used, the vertical distance between the two lines will increase from $2q = p(t) = ct(t')$, which may be vanishingly small, to $2q_0 = p(c) = ct(c')$, which (relative to some theoretical background knowledge) may be much greater.

The central problem so far not solved is the problem of the comparison of the truthlikeness of false theories. I have been unable to show so far that we can approach truth through better and better approximations—through false theories which come nearer and nearer to the truth. I shall show that this is possible for a trivial case: a true theory stating a numerical constant.

---

[1] I must mention here that I disagree in some fundamental points with Harris's paper. More especially, I do not regard the problem of corroboration as a special case of the problem of verisimilitude, and still less as a problem of the agreement of a theory with the experimental data. (This view of Harris's I should be inclined to regard as tinged by inductivism.)

Let us assume that the following theory t is exactly true (though not necessarily the complete truth expressible in our language): $t =$ 'The velocity of light in a vacuum is 3.10^{10} cm/sec.' It will then still be true that the value of this velocity lies somewhere in the ***interval*** between 3.10^{10} cm/sec $\pm 10^8$ cm/sec. Let us call this weaker statement $t \vee i$, where '$\vee i$' denotes the phrase 'or C is equal to a value within a distance of $\pm 10^8$ cm/sec from t'. Let us call $t \vee j$ a similar statement with a greater interval j, so that $t \vee i$ logically entails $t \vee j$. We can then choose $C_i = t \vee i$ or else $C_j = t \vee j$ as our comparison theory; and we can say that, the smaller the interval i (or the stronger the comparison theory), the better does C_i approximate to the exact theory t.

Now let there be two false theories a and b which are both false and both relevant to our problem (the velocity of light, say). Then we can say:

a is nearer to the true theory t than b if and only if there is an approximation theory $t \vee i$ which follows from a but which does not follow from b (while a weaker approximation $t \vee j$ may follow from b).

4 *On Problems Raised by Miller*

David Miller published [1975*a*] the following results, bearing on the problem whether verisimilitude is definable.

(A) If we have two competing false theories, a and b, which predict *more than one numerical constant*, and if the predictions of a are more accurate than those of b, then we can always define new constants such that, with respect to these, the accuracy of the theories is reversed.

(B) A closely similar result is obtained if two theories predict the values of *more than one variable parameter* (say, a time dependent parameter). (See Miller [1975*a*], pp. 170, 180 f.)

I shall first assume the truth of (A), and even of (B). But while I regard Miller's criticism of Tichý's proposal as extremely serious—comparisons of verisimilitude ought to be invariant with respect to translations—I am sceptical about a sceptical interpretation of these further results of Miller's, for the following reasons.

(1) Miller (pp. 185–6) correctly reports me as saying that 'scientific theories' invented 'for the solution of problems are entitled to take as relatively fundamental those parameters that are central to those problems'. What Miller says in reply to this (on p. 186) does not seem to me adequate. He seems to think that, in taking a problem P_1 (and the parameters belonging to it) as fundamental, I somehow have to reject another problem, say P_3, and with it possibly that set of new parameters which would produce Miller's result (B). (This, he rightly shows, would indeed violate one of my own principles: that 'every good theory proffers solutions to problems that it was never designed to solve'.) But I do not reject P_3, nor any other problem that may arise out of the critical discussion of the theory. I have only to remind Miller of my old schema (where P stands for 'problem', TT for 'tentative theory', EE for 'error elimination' by critical discussion and testing).

$$P_1 \rightarrow TT_1 \rightarrow EE_1 \rightarrow P_2 \rightarrow TT_2 \rightarrow EE_2 \rightarrow P_3 \rightarrow etc.$$

There is no particular likelihood that any of Miller's rather *ad hoc* constructed new parameters will turn up as belonging to one of these problems, $P_1, P_2, P_3, \ldots$ But if they do—let us assume they belong to P_3—there is no difficulty. It will just mean that we will demand of TT_3 that it solves not only P_1 and P_2 but also P_3 (with its 'bizarre' set of parameters) better than do TT_1 and TT_2; where 'better' refers to the parameters belonging to each of the problems in question. Thus the fact (which I have admitted for argument's sake) that there are always new problems which may upset the rating of TT_1 and TT_2 presents no new difficulty: if P_3 is such a problem, then TT_3 will have to be better not only with respect to P_3 but also with respect to all preceding problems: this is demanded by the principles ensuring the rationality of the growth of knowledge.[1] (Of course, we may fail to find such a TT_3; but this is an entirely different matter. Also, it is understood that TT_3 may turn out not to be good if confronted with P_4.)

All this shows that the way of approaching truth will partly depend upon the succession of problems; that is to say, upon the history of thought. But this is no more 'backward-looking' than 'forward-looking'.

The 'relativisation' of theories to the problems they are designed to solve is not introduced here *ad hoc*. It is a point which I stressed long ago, and it is needed for solving many problems such as problems concerning the unity and simplicity of theories.

Thus Miller's interesting results (assumed here to be correct) interfere neither with the idea of an approach to truth nor with my methodology (nor, incidentally, with my anti-essentialism). But his results may possibly force us to an historical relativisation of the idea of verisimilitude, in the sense that two historically isolated and different chains of problems *cum* solutions may become comparable with respect to verisimilitude only after the two chains have merged; that is, after we have found theories that solve the problems of both chains better than all their predecessors. (This, incidentally, may throw some new light on the intellectual stimulus of culture clash.)

(2) But we must also be clear that the approach of numerical test statements to truth is far from decisive for the approach of theories to truth. Only in cases such as statements of constants, such as c or h, is the approach to the true numerical value the same as the approach to truth. In all other cases, there are qualitative or topological conjectures involved, and these are not subject to Miller's methods (A) and (B). Moreover, there are conjectures about structure, which are most important.

(3) It may be (this is very tentative) that it is a reply to Miller's argument (A) concerning the combination of two numerical constants that, in testing the value of a constant, we avoid testing the value of two constants together. (If there is more than one constant involved, we treat only one as problematic and the others as 'background knowledge'.) This may be due to the scientist's intuitive (or instinctive) avoidance of the situation described by Miller.

(4) So far I have assumed the validity of Miller's arguments (A) and (B). Yet I do not see how these arguments can be squared with the practice of certain computerised guided missiles (or space missions). Here the target may be described by more than one parameter (perhaps by a set of positional parameters).

---

[1] See my [1975], section VIII.

The missile is guided by feedback, that is to say, by the relative evaluation of tentative predictions, precisely of the kind demanded by the theory of verisimilitude. If the computer can transform the parameters into an equivalent set for which the evaluations of the predictions are reversed, how can it ever succeed in its task? And if it *does* succeed, by the method of comparing and evaluating a sequence of false predictions, why cannot we do the same? This point seems to me in urgent need of clarification.

5 *Concluding Remarks*

As I have often emphasized, a statement like 'The theory *a* is nearer to the truth than the competing theory *b*' is never demonstrable, but may be nevertheless asserted as a conjecture, strongly arguable for or against on the basis of (1) a comparison of the logical strength of the two theories and (2) a comparison of the state of their critical discussion, including the severity of tests which they have passed or failed. ((2) can also be described as a comparison of their degree of corroboration.)

It seems to me that, on the basis of such comparisons, we can give good reasons in support of the conjecture that Einstein's theory of gravitation is nearer to the truth than Newton's.

(1) Einstein's theory seems to have greater content, since it seems to solve all relevant problems which Newton's can solve (because Newton's theory can be obtained from it in approximation). On the other hand, there are several great problems on which Newton's theory has nothing to say while Einstein's has. (Example the redshift of light emitted in a strong gravitational field.)

(2) The tests appear to show that the numerical results of Einstein's theory are in a few fields nearer to the target than those of Newton's theory (Mercury perihelium movement; light rays grazing the surface of the sun).

This analysis, which I have sketched before ([1954]; see [1959] p. 401) for degree of corroboration, has been often criticised. I still think that it is essentially valid.

To conclude I wish to refer to the case of non-numerical target theories, such as in the analysis of what I have called social situations (or situational logic). In both cases we wish to approximate the objective truth, a true description of the facts. But while in the case of numerical theories the approximation appears to be subject to a simple straightforward logical analysis, in the non-numerical case this kind of analysis seems to be difficult, if not impossible.

The very idea of a true target theory seems difficult, for the target may not be theoretical and abstract but rather a concrete situation (for example, in history), of which we have only a very partial description. The concrete situation itself may be infinitely complex; and so an abstracting or selecting procedure—an interpretation—seems needed which cannot avoid a chosen subjective selecting point of view. This does not mean, however, that our analysis, as far as it goes, cannot be objectively true or false, or that it cannot be objectively tested. It only means that new, relevant, and previously unsuspected aspects may always be discovered in the situation, however far we may have driven our analysis.[1]

Of course, it is possible that we may find two competing theories, *a* and *b*, which cannot be compared in this way or in other ways. For example, we may have several exact numerical target theories t_1, t_2, . . ., and *a* may be a better approximation with respect to the target t_1, while *b* is a better approximation with respect to the target t_2. I do not think that this situation creates a serious problem: we shall, in this case, aim at constructing a theory with the advantages of *a* and *b*, without their disadvantages. Thus we shall still follow the idea of obtaining a better approximation to truth.

---

[1] For further details see my paper 'Models, Instruments and Truth', forthcoming.

KARL POPPER
Penn, Buckinghamshire

REFERENCES

Bernstein, B. A. [1936]: 'Postulates for Boolean Algebra Involving the Operation of Complete Disjunction', *Annals of Mathematics*, **37**, pp. 317–25.

Fréchet, M. [1930]: 'Sur la convergence "en probabilité"', *Metron*, **8**, pp. 3–50.

Fréchet, M. [1931]: 'Nouvelles expressions de la "distance" de deux variables aléatoires et de la "distance" de deux fonctions measurables', *Annales de la société polonaise de Mathématique*, **9**, Année 1930, pp. 45–9.

Harris, J. H. [1974]: 'Popper's Definitions of Verisimilitude', *British Journal for the Philosophy of Science*, **25**, pp. 160–6.

Mazurkiewicz, S. [1933]: 'Zur Axiomatik der Wahrscheinlichkeitsrechnung', *Comptes rendus des séances de la société des sciences et des lettres de Varsovie*, **25**, Classe III, Année 1932, pp. 1–4.

Mazurkiewicz, S. [1934]: 'Ueber die Grundlagen der Wahrscheinlichkeitsrechnung I', *Monatshefte für Mathematik und Physik*, **41**, pp. 343–52.

Miller, D. [1972]: 'The Truth-likeness of Truthlikeness', *Analysis*, **33**, pp. 50–5.

Miller, D. [1974*a*]: 'Popper's Qualitative Theory of Verisimilitude', *British Journal for the Philosophy of Science*, **25**, pp. 166–77.

Miller, D. [1974*b*]: 'On the Comparison of False Theories by their Bases', *British Journal for the Philosophy of Science*, **25**, pp. 178–88.

Miller, D. [1975*a*]: 'The Accuracy of Predictions', *Synthese*, **30**, pp. 159–91.

Miller, D. [1975*b*]: 'The Accuracy of Predictions: A Reply', *Synthese*, **30**, pp. 207–19.

Nikodym, O. [1930]: 'Sur une généralisation des intégrales de M. J. Radon', *Fundamenta Mathematica*, **15**, pp. 131–79.

Popper, K. [1938]: 'A Set of Independent Axioms for Probability', *Mind*, N.S., **47**, pp. 275–7. (Reprinted in Popper [1959], pp. 318–22.)

Popper, K. [1945]: *The Open Society and its Enemies*. (10th impression 1974.)

Popper, K. [1954]: 'Degree of Confirmation', *British Journal for the Philosophy of Science*, **5**, pp. 143–9. (Reprinted in Popper [1959], pp. 395–402.

Popper, K. [1959]: *The Logic of Scientific Discovery*. (8th impression 1975.)

Popper, K. [1963]: *Conjectures and Refutations*. (5th impression 1974.)

Popper, K. [1966]: 'A Theorem on Truth-Content', in P. Feyerabend and G. Maxwell (*eds.*): *Mind, Matter and Method, Essays in Philosophy and Science in Honor of Herbert Feigl*, pp. 343–53.

Popper, K. [1972]: *Objective Knowledge*. (4th impression 1975.)

Popper, K. [1975]: 'The Rationality of Scientific Revolutions', in R. Harré (*ed.*): *Problems of Scientific Revolution: Progress and Obstacles to Progress in the Sciences*, pp. 72–101.

Tarski, A. [1935], [1936*a*]: 'Grundzüge des Systemenkalküls I, II', English translation in Tarski [1956], pp. 342–83.

Tarski, A. [1936*b*] (in Polish, 1931): 'Der Wahrheitsbegriff in den formalisierten Sprachen'. English translation in Tarski [1956], pp. 152–278.

Tarski, A. [1956]: *Logic, Semantics, Meta-Mathematics. Papers from 1923 to 1938*, translated by J. H. Woodger.

Tichý, P. [1974]: 'On Popper's Definitions of Verisimilitude', *British Journal for the Philosophy of Science*, **25**, pp. 155–60.

Tichý, P. [1976]: 'Verisimilitude Redefined', *British Journal for the Philosophy of Science*, **27**, pp. 25–42.

COMMENTS ON LANDÉ

Landé's article in the last issue (pp. 43–50) and my review of his recent book in the previous issue (pp. 355–8) were each written without knowledge of the other, so it seems appropriate to clarify where our present disagreement lies.

Landé evidently now retracts (p. 44 and p. 48) what I called claim (I) in my review. This seems to me something of a *volte face* when compared with all his previous publications. His article is concerned solely with defending what I called claim (II) in my review.

My criticisms of claim (II) in my review still seem to me to be apposite. In particular it seems to me that there is a fundamental sleight-of-hand underlying Landé's construction: we expect observable quantities only to depend on position differences and on momentum differences (for we reject absolute positions and absolute velocities), Landé wishes us to slide from this into accepting his conditions on matrix elements. But (*i*) as Aharanov has emphasised (Aharanov and Petersen [1971], Aharanov and Susskind [1967]), quantum mechanical position and momentum operators do not correspond to absolute position and momentum but to position and momentum relative to a specifiable frame of reference; therefore the quantum mechanical formalism already takes into account relativity of positions and velocities even without our imposing Landé's restrictions on matrix elements. (*ii*) If Landé's philosophical slide were justified then not merely the matrix element of any observable which was a function of position only, but the matrix element of any observable whatsoever, ought to depend only on the momentum difference between two states of momentum in the x-direction. But this more general conclusion would make complete nonsense of Landé's putative derivation of the commutation relations. Taking this criticism together with those in my review, it seems to me difficult to attach any significance to Landé's alleged new justification of the quantum theory.

JON DORLING
Chelsea College,
University of London

REFERENCES

AHARANOV, Y. and PETERSEN, A. [1971]: 'Definability and Measurement in Quantum Theory' in E. Bastin (*ed.*) *Quantum Theory and Beyond* (Cambridge, 1971), pp. 135–9.

AHARANOV, Y. and SUSSKIND, L. [1967]: 'Charge Superselection Rule', *Physical Review*, 155, pp. 1428–31.

THE APPLICABILITY OF BAYESIAN CONVERGENCE-OF-OPINION THEOREMS TO THE CASE OF ACTUAL SCIENTIFIC INFERENCE

Section II of Dorling [1975] attempted to formulate and prove a convergence-of-opinion theorem appropriate to Bayesian reconstructions of ordinary scientific inference. The proof depended on providing an accurate formal rendering of what it is for evidence *on average* to favour one hypothesis h rather than another hypothesis g (but not to an increasingly negligible extent). Edgington pointed out to Dorling in correspondence that his formal rendering of this is inadequate and leads to paradox in addition to generating a formal fallacy at a later point in the proof.

Dorling's formulation depended on requiring that a certain average likelihood ratio be greater than 1. However the difficulty is that if two likelihood ratios are reciprocals of each other (say $\frac{1}{3}$ and 3) they ought intuitively to average to 1, whereas Dorling's formulation required arithmetical averaging ($\frac{1}{3}$ and 3 averaging to $1\frac{2}{3}$), with the result that according to his criterion the same evidence could both favour h as against g, on average, and favour g against h, on average. This is plainly paradoxical and is also inconsistent with the conclusion of his theorem.

An intuitively correct formal rendering here seems to require that we consider geometrical rather than arithmetical averages of likelihood ratios. Now the geometrical mean of a set of n likelihood ratios is the nth root of their product. It is this function then, of Dorling's likelihood ratios (rather than their arithmetical average) which should be required to be greater than $(1+\epsilon)$ 'for any sufficiently large n'. But then it immediately follows that the product of these ratios must be greater than $(1+\epsilon)^n$. This is precisely the inequality we need to restore the final step of Dorling's proof, a step which Edgington noted was fallacious on the basis of his original inequality.

Subject to the above modifications it seems that Dorling's proof provides a satisfactory answer to Hesse's criticism of Savage's original result as depending on too restrictive assumptions for ordinary cases of scientific inference.

JON DORLING
Chelsea College
DOROTHY EDGINGTON
Birkbeck College

REFERENCE

DORLING, J. [1975]: Review of M. B. Hesse [1974]: *The Structure of Scientific Inference*, *British Journal for the Philosophy of Science*, **26**, pp. 61–71.

NEWCOMB'S 'PARADOX'

In a recent article ([1974]) G. Schlesinger has argued that Newcomb's paradox provides an argument for supposing that virtually no amount of inductive evidence could support the claim that a given theory, device, or agent is capable of predicting human choices with a better than chance expectation of success. Roughly put, his argument claims that the existence of such a predictor makes it possible for a certain game to be played (the Newcomb game) in which two contradictory courses of action are to be recommended as exclusively rational courses to follow. (We shall take the Newcomb game to be described as follows. There are two boxes, I and II, such that $1,000 is placed in I. Before the player chooses between simply taking box II or taking both boxes I and II the predictor predicts his choice and when the player leaves the area of the boxes $1,000,000 is placed in box II if and only if the prediction is that the player will take only box II. The boxes are then covered and the player returns to make his choice.)

If Bar-Hillel and Margalit ([1972]) are correct, and Schlesinger agrees that they are, inductive support for the claim that the predictor is likely to be correct with probability *r* amounts to assuming that the probability that a million dollars is in box II is *r* if that box is chosen and $1-r$ if both are taken. Thus, inductive support for the claim that the predictor is always correct yields the conclusion that if the player wishes to maximise his expected utility, then he should choose only box II. (Indeed one need not even suppose the predictor is correct with probability 1 to arrive at the conclusion one ought to take box II only.) On the other hand Schlesinger gives what he takes to be an equally sound argument for the rationality of choosing both boxes. Thus he concludes that we must suppose that almost no amount of any inductive evidence could support the belief that the predictor was capable of perfect prediction.

Schlesinger's argument for choosing both boxes runs as follows. First we are to imagine what a 'perfect well wisher . . . who always advises [the player] to do what is in [his] best interest' and who is 'perfectly intelligent' would wish the player to do, if he were able to see the contents of both boxes. Schlesinger points out that whether such a well wisher sees simply $1,000 or $1,001,000 he would wish the player to choose both boxes and thereby maximise his gain. He then argues that since the well wisher always wishes (by definition) the player to do what is in his own best interest, taking both boxes must be in the best interest of the player. From this Schlesinger concludes that the rational course for the player is to take both boxes. Thus the contradiction and Schlesinger's conclusion about predictability in principle.

A little analysis, however, leads to the conclusion that Schlesinger's argument is fundamentally unsound. There are three major points in the argument. First, there is the claim that the well wisher would wish the player to take both boxes. Second, there is the claim that whatever the well wisher wishes the player to do is what is in the player's best interest. (This is supposed to be analytic.) And third, Schlesinger supposes that what is in the player's best interest in what is rational for him to do. The problem in Schlesinger's argument arises because it is so easy to conflate the distinction between what is rational to do and what will produce the best results. Indeed the three claims above are plausible only because the term 'best interest' is used in a fashion which hides the confusion between what is rational and what produces the greatest gain.

Let us consider a simple game in which a player is presented with two boxes, *A* and *B*. Let us also suppose that there is to be either a million dollars in box *A* and nothing in *B* or else a million in *B* and nothing in *A*. Let us further suppose that there is a great deal of evidence to support the player's belief that the probability is .9 that the money is in box *A* and .1 that it is in box *B*. Clearly the player should, if he wishes to make the rational choice, choose box *A*. Let us now suppose that in fact the money is in box *B* and that a 'perfect well wisher' can see this to be the case. Now if 'well wisher' is defined as Schlesinger defines it, it seems that we can easily argue that the well wisher desires the player to choose box *B*. For, after all, does he not desire the player to do what is in the player's own best interest? Surely it is in the player's best interest to receive the million as opposed to nothing. Thus the well wisher wishes the player to choose box *B* even though this is the irrational choice for him to make.

The point is that we can and must distinguish between (*a*) what I should do in the sense of what rational policy I should follow and (*b*) what I should do in the sense of what will in fact attain the most desirable result for me. That Schlesinger does not make this distinction can be seen in his view that it is plain that the well wisher would wish the player to take both boxes and that this action is in the player's best interest and hence is the rational choice. If one asks a friend for advice and he advises one in a certain way, then usually it follows that if one's friend is generous and intelligent it would be rational to follow his advice. But the fact that one's friend wishes one to act in some manner and wishes the best for one *does not entail that it would be rational to do what he wishes one to do.* (The simple game above refutes this contention.) The only way in which Schlesinger's first claim (that the well wisher wishes the player to take both boxes) would be plausible is if we suppose the well wisher to wish the player to do what would bring him the greatest gain—for the well wisher sees that the player will gain more by taking both boxes than by taking only box II. But then it is not evident that choosing both boxes is rational and Schlesinger's third claim needs further argument. If, on the other hand, we suppose that 'best interest' is to mean the same as 'rational', then Schlesinger's third claim is analytic but his first unsupported. For if a perfect well wisher is someone who always wishes the player to do what is rational, we cannot assume he will wish the player to take both boxes despite what he sees in them.

If this last claim seems strange, consider the following possibility. The well wisher (wishing the player to make the rational choice) sees that box II is empty. He says the following to himself. 'The fool, I wish he would choose only box II although of course he won't. Were he to choose simply box II I would not now be seeing a paltry thousand in box I but also a million in box II.' No doubt the tenses in these sentences seem strange for we are not used to thinking that present circumstances can be dependent upon future choices but the reasoning above does seem comprehensible and even plausible. At any rate, unless it is already established that the rational choice for the player is to choose both boxes one cannot conclude that the well wisher, defined as he who wishes the player to make the rational choice, will wish the player to choose both.

If we compare the Newcomb game to the simple choice game above which refutes the principle of doing what a friend would have one do as the rational course, we can see that there is a difference between the two. In the simple choice game it is not rational to do what a well wishing (in the sense of wishing the player to do what will bring him the greatest gain) friend would have you do because one does not know what that is. However it is quite rational to do what such a friend wishes if one knows what his advice is. Thus although Schlesinger fails to give an immediately convincing argument because he fails to realise that his tacit assumptions need support, one might ask whether or not his argument can be justified through further analysis.

So the situation is this. The player of the Newcomb game has returned and is facing the two closed boxes. Across the room sits a friend of his who is very intelligent and who wishes the player to make the choice that will bring the player his greatest gain. Furthermore the player knows that the friend wishes him to take both boxes. Although he cannot appeal to any obviously analytic principle to establish the rational course, as Schlesinger mistakenly does, we can ask the same question he does. Why would the player be irrational to follow his friend's advice? The answer seems to be this. Knowing what the player's friend advises does not give the player any information about whether box II is empty or contains a million dollars. But, there is inductive evidence to support the contention that the decision the player makes affects the probability that there is a million dollars in box II. Thus the player reasons as follows. Although I know that my friend wishes me to take both boxes I do not know why. Does he wish me to take both because box II is empty and taking both will bring me at least a thousand or because box II contains a million and he wishes me to have $1,001,000? Since I will gain more if I choose simply box II and thereby insure the gain of a million than if I choose both and gain only a thousand I will disregard his advice. The important distinction between the situation in the Newcomb game and other games of choice is that knowing what choice one's friend would have one make does not provide any information about what the situation is. This is quite in contrast to the situation where one's friend's advice literally tells one what the distribution of money is.

What is peculiar about the Newcomb game is that it provides a situation in which one has to think that the probabilities of present (or past) situations are determined by future choices. However, we do not believe that there is any *a priori* argument to show that one cannot have inductive evidence for believing that a choice, as yet unmade, could affect the probability that some proposition or other is now true. At any rate, the Newcomb game does not appear to show this and unless some other argument can be given with more force than Schlesinger's we must suppose that there is only one rational choice in the Newcomb game, namely choose box II alone.

T. M. BENDITT
University of Southern California

DAVID J. ROSS
Duke University

REFERENCES

BAR-HILLEL, Y. and MARGALIT, A. [1972]: 'Newcomb's Paradox Revisited', *British Journal for the Philosophy of Science*, **23**, pp. 295–304.

SCHLESINGER, G. [1974]: 'The Unpredictability of Free Choices', *British Journal for the Philosophy of Science*, **25**, pp. 209–21.

Brit. J. Phil. Sci. **27** (1976), 165–176 *Printed in Great Britain*

Review Articles

ELKANA ON HELMHOLTZ AND THE CONSERVATION OF ENERGY*

Elkana's new book is intended as a contribution both to the history and to the philosophy of science. The author's central historical concern is the discovery of the law of conservation of energy which he takes as having been made by Helmholtz and reported in his classic paper '*Die Erhaltung der Kraft*' of 1847. Elkana offers historical analyses of the intellectual traditions—in mechanics (ch. II), in heat and energy (ch. IV) and in physiology (ch. V)—which played a role in the discovery of the law, and of the social and institutional factors which he claims were necessary conditions for the discovery of the law. The volume is intended as a case-study in Elkana's distinctive methodology, 'the image of science'.[1]

The volume contains many interesting and suggestive asides but it seems to me that at least three of Elkana's central historical claims are false *viz.*:

(*i*) that Helmholtz proved 'correctly and generally the law of the conservation of energy' (p. 16);
(*ii*) that Helmholtz had a vague and undefined concept of '*kraft*' (which covered what we now consider to be the distinct notions of force and energy) and that had he not held such a vague notion he would not have arrived at a conservation of energy principle;
(*iii*) that the rôle of other scientists in the discovery of the principle was relatively small.

In the remainder of this review I shall attempt to substantiate my claim that (*i*), (*ii*) and (*iii*) are false and to examine some subsidiary claims argued by the author (sections 1–6). One cannot but conclude that Elkana was led into error by his philosophy. In particular, as will be argued below (section 7), his philosophical claims based upon the notions of 'concept in flux' and 'image of science' seem confused and often misleading.

1 Elkana claims (p. 16) that Helmholtz proved 'correctly and generally the law of the conservation of energy', in his 1847 paper. Further he remarks of current appraisals of Helmholtz's achievement that, 'It is admitted by all, that the man who formulated the principle for the first time mathematically, in all its generality, was Hermann von Helmholtz . . .' (p. 9). Nowhere in the volume,

---

*Review of Elkana, Y. [1974]: *The Discovery of the Conservation of Energy*; London: Hutchinson Educational. £4.00. Pp. x+213. This volume contains as an appendix the author's earlier paper (Elkana [1970*a*]).

I should like to thank Gregory Currie, Jon Dorling, David Lavis, Alan Musgrave, Peter Urbach and Elie Zahar for helpful criticism of earlier drafts of this review.

[1] A second recently published case-study concerns Boltzmann's scientific research programme and its rivals (Elkana [1975]). A general characterisation of this methodology can be found in Elkana [1970*b*] and his [1971].

however, does the author indicate that both of the major theorems given in Helmholtz's paper are in fact false. To see this it is necessary to examine Helmholtz's argument in detail. In the first section of his paper he makes two assumptions. The first is that 'all actions in nature can be ultimately referred to attractive or repulsive forces, the intensity of which depends solely upon the distances between the points by which the forces are exerted'.[1] The second assumption amounts to the exclusion of perpetual motion, in the sense of the impossibility, 'by any combination whatever of natural bodies to derive an unlimited amount of work force'.[2] He takes the mathematical form of the latter assumption to be the well-known theorem of mechanics which asserts the conservation of *vis viva*.[3] Finally he asks under what conditions this conservation principle holds. His answer is that it holds when, and only when, *all* the forces involved are central ones. In the second section of the paper Helmholtz seeks to give 'the law for the cases where central forces act a still more general expression'. He argues that the difference between the *vis viva* of a particle at distances r and R from a fixed mass point is equal to the quantity $\int_r^R \phi \, dr$ or the sum of the tensions [*Sümme der Spannkräfte*] *i.e.*, 'the sum of the intensities of the forces which act at all distances between r and R'.[4] Thus if q and Q correspond to the velocities at r and R then,

$$\tfrac{1}{2}mQ^2 - \tfrac{1}{2}mq^2 = -\int_r^R \phi \, dr.^5$$

In general for a system of n-mass points $(m_1, \ldots., m_n)$ among which only central forces ϕ_{ij} exist, the law assumes the form,

$$-\sum_{i,j}\left(\int_{r_{i,j}}^{R_{i,j}} \phi_{i,j} \, dr_{i,j}\right) = \sum_i \left(\tfrac{1}{2}m_i Q_i^2 - \tfrac{1}{2} m_i q_i^2\right).$$

That is in general 'the sum of the tensional forces and the *vires vivae* is always constant'.[6] This is Helmholtz's law of the '*Erhaltung der Kraft*'. He concludes the theoretical part of his paper by emphasising that this law holds when and only when central forces act. He writes, 'If, however, other than central forces had existence, rigid combinations of natural bodies might be formed which could move of themselves without needing any relation whatever to other bodies', and that, 'If natural bodies [were] possessed of forces which depend upon a time and velocity . . . then combinations of such bodies would be possible in which force might be either lost or gained *ad infinitum*'[7].

Now considering Helmholtz's claim that conservation of *vis viva* holds if and only if all the forces are central, a simple counterexample to the 'only if' claim is provided by a uniform field of force ($\vec{F}$ = constant). Of much greater historical importance are the counterexamples to Helmholtz's claim that the law of the '*Erhaltung der Kraft*' holds when and only when the forces are central. These counterexamples are provided by electromagnetic theories in which the forces are dependent upon the velocity. Elkana gives no account of the major criticisms of this, Helmholtz's central claim made by Clausius, Weber and ultimately Maxwell, even though he considers reactions well into the 1870s. Clausius's criticism was directed precisely at Helmholtz's claim for the necessity of central forces.[1] He pointed out that if many forces act on a mass point in differing directions the net force need not be central yet the system will still obey the conservation law. In Weber's action-at-a-distance theory the magnitude of the attraction and repulsion depend both upon the relative velocity of approach and upon the acceleration. Thus if Helmholtz's theorem were correct, Weber's theory must be inconsistent with the conservation of energy and therefore must be rejected.[2] Weber replied to the effect that conservation may have to be abandoned in electromagnetism.[3] Maxwell however initially accepted Helmholtz's result and rejected Weber's theory. He remarks:

---

[1] All forces are then of the form $\vec{F} = M\,\phi(r)\,\vec{r}$ (Helmholtz [1847], p. 114). Forces which depend explicitly upon the time and/or the velocity do not occur in nature according to Helmholtz.

[2] *Ibid.* p. 114. The aim of the paper is essentially to prove that these two assumptions are equivalent.

[3] In the case of the conservation of *vis viva*, the quantity of work performed is expressed as the equivalent of a weight raised to a given height (mgh). The law is given by Helmholtz as $\tfrac{1}{2}mv^2 = mgh$ (*ibid.* p. 119).

[4] *Ibid.*, pp. 119–20.

[5] *Ibid.*, p. 120.

[6] *Ibid.*, p. 124.

[7] *Ibid.*, p. 131.

> The mechanical difficulties, however, which are involved in the assumption of particles acting at a distance with forces which depend on their velocities are such as to prevent me from considering this [Weber and Newmann] theory as an ultimate one, though it may have been, and may yet be useful in leading to the coordination of phenomena.
>
> I have therefore preferred to seek an explanation of the fact in another direction, by supposing them to be produced by actions which go on in the surrounding medium as well as in the excited bodies, and endeavouring to explain the action between distant bodies without assuming the existence of forces capable of acting directly at sensible distances.[4]

By 1873, however, Maxwell had realised the error in Helmholtz's argument and abandoned it as a criticism of Weber's theory. He writes in the *Treatise*, '. . . a law of electrical action involving the velocity of particles, has sometimes been supposed to be inconsistent with the principle of the conservation of energy. The formula of Gauss is inconsistent with this principle and must therefore be abandoned, as it leads to the conclusion that energy might be indefinitely generated in a finite system by physical means. This objection does not apply to the formula of Weber . . .'.[5] My claim is here that Elkana in placing so great a stress on only one aspect of Helmholtz's paper (the assertion of an energy conservation law), takes Helmholtz's results out of context (the central claimed equivalence of the conservation law with the universality of central forces), a procedure which results in an incorrect appraisal of both Helmholtz's actual result and its significance in relation to the rest of contemporary physics.

---

[1] *Cf.* Clausius [1853]. Heimann in his [1974*b*] gives a very thorough account of criticisms of Helmholtz's paper including those of Clausius, Weber and Maxwell (Heimann [1974*b*], pp. 232–8). Dorling in his [1970] clearly points out Helmholtz's error and Maxwell's treatment of it in the *Treatise* (Dorling [1970], pp. 262–65). Helmholtz's replies to Clausius were not very satisfactory. Helmholtz stated that he was not concerned with mathematically possible counterexamples but what was physically the case and that Clausius had not clearly distinguished between 'points' and 'corporeal elements'. *Cf.* Heimann [1974*b*], pp. 234–5.

[2] Helmholtz [1847], pp. 156–60. *Cf.* Heimann [1974*b*], pp. 234–5.

[3] Weber [1869], [1878].

[4] Maxwell [1865], p. 527.

[5] Quoted by Dorling (Dorling [1970], p. 243).

2 A second historical thesis which Elkana argues is that, 'at the time (1847) the two concepts [energy and force] were not at all clearly defined separate entities, but rather different disguises of a vaguely defined '*Kraft*' which was being conserved' (p. 16). Helmholtz's '*Kraft*' is according to the author an important illustration of a 'concept in flux', and he argues that 'the confusion between "force and energy" in the works of Helmholtz and some of his contemporaries was not only a verbal one . . . but rather a *necessary* prerequisite for the final clarification of concepts' (p. 9).[1] This section will be devoted to an analysis of this claim. Now Elkana is not arguing for the triviality that before concepts are clarified they must necessarily have been vague. His claim is that only a vague, undefined, undifferentiated '*Kraft*' could have been thought subject to a conservation law: 'Only an undefined entity *could have been* the subject of a general belief in the principles of conservation in nature,' (p. 9, my italics). The claim for deep conceptual ambiguity in Helmholtz's use of '*Kraft*' is supported by six pages of textual analysis (pp. 132–8). Elkana is not arguing here that Helmholtz uses the term '*Kraft*' in several different ways, the meaning of the term in a given case being clear from the mathematical or discursive context, but that there is a single entity '*Kraft*' which is thought of as conserved. This, if it means anything at all, means that Helmholtz did not distinguish between force and energy terms.[2] Yet nowhere in Helmholtz's paper is there any suggestion that forces in the sense of the Newtonian $\vec{F}=d(m\vec{v})/dt$ are conserved. In the first section of his paper (referred to *above*, p. 166) there is no confusion between force terms and energy terms. The force terms are all of the form $\vec{F} = m\phi(r)\vec{r}$ and the energy terms $\frac{1}{2}mq^2$, for q the velocity of a mass point. In the second section of the paper, it is clear that Helmholtz had two distinct classes of entities in mind. First the tensions (or forces ϕ) and second their *sum* and the $\frac{1}{2}mv$'s and their sum. Presumably if Helmholtz had had only a single conserved entity in mind, as the author claims, then one would expect to find some reflection of this in the mathematical form of the argument. But the forces are never, for example, compared with the square of the velocities; it is always the sum $\int_r^R \phi\, dr$ which is so compared. It is worth quoting directly from Helmholtz on the meaning of $\int_r^R \phi\, dr$ since he makes his intention quite clear:

> To understand the import of the quantity $\int_r^R \phi\, dr$, let us suppose the intensities of ϕ which belong to different points of the connecting line ma erected as ordinates at these points, then the above quantity would denote the superficial content of the space enclosed between the two ordinates r and R. As this surface may be regarded as the sum of the infinite number

---

[1] My italics. The same claim is argued extensively in Elkana [1970*b*].

[2] This interpretation of Elkana's thesis is also adopted by Heimann. *Cf.* especially his [1974*b*], p. 207, n. 10 and p. 209, n. 14.

> of ordinates which lie between r and R, it therefore represents the sum of the intensities of the forces which act at all distances between R and r. Calling the forces which tend to move the point m, before the motion has actually taken place, *tensions*, in opposition to that which in mechanics is named *vis viva*, then the quantity $\int_r^R \phi\, dr$ would then be *the sum of the tensions* between the distances R and r, and the above law would be thus expressed: The increase of *vis viva* of a material point during its motion under the influence of a central force is equal to the sum of the tensions which correspond to the alteration of its distance.[1]

It seems to me that force terms (ϕ) and energy terms ($\int \phi\, dr$, $\frac{1}{2}mv^2$) are here quite distinct. It is certainly correct that Helmholtz employs the term '*kraft*' in different contexts: *lebendige kraft* (*vis viva*), *bewegendekraft* (as a translation of the French '*puissance motrice*'), *spannkräfte* (tensional forces), and *arbeitskraft* (work force). But the context, and the translations of these terms into the mathematical argument show that they have clear and distinct interpretations.[2] Since Helmholtz's major premise is the impossibility of a perpetual motion machine, it seems likely that his conservation of 'force' principle was meant to indicate not only the law $\frac{1}{2}mQ^2 - \frac{1}{2}mq^2 = -\int_r^R \phi\, dr$ but the indestructibility and transformability of *all* agents whatsoever which could effect an action or perform work.[3] This latter very general sense of the law of '*Erhaltung der Kraft*' reflects the origins of Helmholtz's interest in the conservation of working power in the problem of animal heat. There is some ambiguity here in the sense of '*Kraft*' as 'effecting an action'. But to say as Elkana does that 'the very ambiguity in the word '*Kraft*' is indispensable for the comprehensibility of the work' (p. 17), given the form of Helmholtz's argument is false.

Elkana summarises the programme executed by Helmholtz in his paper as follows (p. 129):

> If we collect these premises systematically the argument looks as follows:
> (i) Newtonian 'force' is a fundamental concept in mechanics.
> (ii) Physics is reducible to mechanics.
> (iii) The fundamental concept in physiology is 'force of life'; physiology is reducible to physics, *i.e.*, to mechanics.
> (iv) There is a basic entity in Nature which is being conserved.
> Conclusion: the basic entity which is being conserved must be 'Kraft'. But for the final formulation we have to add another premise:
> (v) The Lagrangian formulation of mechanics is equivalent to the Newtonian formulation mathematically and conceptually: The Lagrangian formulation has as its fundamental entity 'kinetic energy plus potential energy'; this fundamental sum is being conserved.

---

[1] *Op. cit.*, p. 122. It is interesting to note that Helmholtz employs here the concept of the definite integral as the sum of the uncountably many ordinates between the limits of integration.

[2] A detailed examination of the different uses of '*Kraft*' in Helmholtz's paper in their contextual setting is given by Heimann in his [1974*b*], especially pp. 213–6.

[3] *Cf.* Heimann [1947*b*], p. 207.

> Conclusion: The basic entity 'Kraft' which is being conserved in Nature must be equivalent in dimension and form to mechanical energy. This is the generalised conservation of energy principle.

Now if one examines Elkana's reconstruction it is clear that if the first conclusion therein is to have any semblance of validity it must be rendered: 'The basic entity which is being conserved must be "force"', *i.e.*, Newtonian force. But this is never argued by Helmholtz. Given the form of Helmholtz's argument it is clear that force terms and energy terms are kept distinct, they are not subsumed under a single vague concept '*Kraft*'. If this is correct then it is clear that Elkana is wrong in concluding that *only* an undifferentiated entity '*Kraft*' could have been thought subject to a conservation law.[1]

3 Let us now turn from Elkana's analysis of the content of Helmholtz's classic paper to his account of the intellectual and social background to it. As remarked above, Helmholtz's idea, that the ability of natural agents or combinations of bodies (machines) to perform work is conserved, originated in the problem of animal heat. Elkana treats this important problem extensively. In three informative sections (pp. 97–113) he examines the tradition of vitalism in the eighteenth century (pp. 98–103), Liebig's theory that respiration is the source of animal heat (pp. 103–7), and Müller's work on the physiology of the senses and the nervous system (pp. 107–9). Elkana highlights an important problem here: although Liebig in his [1842] claimed that chemical forces (those derivable from the ordinary laws of chemistry) are generated by respiration and that these alone are responsible for animal heat and bodily exertion, he also asserted that a 'vital force' unique to living things must be postulated in order to explain the growth and development of biological individuals. Thus Liebig's theory is at once 'vitalist' and 'reductionist'. Helmholtz, commenting on this theory, made two claims: first that Liebig's theory presupposes a principle of 'the constancy of the force equivalents by the excitation of one force by another',[2] and second that it remained an open question whether Liebig's 'vital force' was some combination of natural forces or a self-perpetuating force unique to life.

Now Elkana takes this as a direct indication that Helmholtz's concepts were 'so vague and in a "state of flux", that he could even accommodate in his conceptual framework some "vital forces", and believe that as such they could be reduced to physico-chemical terms' (p. 107). But here again, the "concepts in flux" are Elkana's, not Helmholtz's, for Helmholtz makes a sharp distinction between a 'vital force' which is some concatenation, given suitable initial conditions, of natural forces (a view consistent with Liebig's), and a 'vital force' as a 'self-perpetuating unique force'. In the first case the 'vital force' would be subject to a conservation-interconvertibility law, while in the second case, being self-perpetuating, it would not. Far from being confused Helmholtz makes an important and fruitful distinction.

---

[1] Helmholtz possessed a sharply defined potential energy concept in the '*Sümme der Spannkräfte*' (indeed the major portion of Helmholtz's paper is devoted to computing this quantity in various applications to mechanics, heat, electricity and magnetism). It was a restricted and cumbersome one, by virtue of the difficulty of calculating the integral in specific applications, compared with the generic concept of the thermodynamic energy as a function of state introduced into thermodynamics by Clausius. *Cf.* Clausius [1854], pp. 112–5.

[2] Quoted in Heimann [1947*b*], p. 210.

4 Elkana often repeats the claim (pp. 12, 129, 158, 169, 187) about the origin of Helmholtz's ideas on conservation that he was strongly committed to an *a priori* belief in conservation principles in nature. Elkana admits however, that, 'There is no direct evidence at my disposal which could carry my point. There are no sweeping Helmholtzian statements about his metaphysical commitments, or about his pre-scientific beliefs' (p. 158). However, the opening section of the paper contains a clearly articulated highly metaphysical thesis: 'we discover the problem of physical natural science to be, to refer natural phenomena back to unchangeable attractive and repulsive forces, whose intensity depends solely upon the distance. The solvability of this problem is the condition of the complete comprehensibility of nature'.[1] The commitment here is quite clear. It is to the universitality of central forces,[2] not to an *a priori* belief that something is conserved. Elkana states that (p. 129), 'Helmholtz was very much committed *a priori* to two fundamental beliefs', these were according to him '(*a*) that all phenomena in physics were reducible to mechanical processes and (*b*) that there must be some basic entity in nature which is being conserved'. There is no direct evidence for (*b*). Further, from an *a priori* viewpoint, the strong evidence for (*a*) argues against (*b*), for Newtonian forces are not conserved, as was well known. The strongly Kantian flavour of Helmholtz's remarks (quoted *above*) suggests a powerful Kantian influence. Here he re-echoes Kant's own avowed metaphysical underpinning of the Newtonian principles of central forces and action at a distance as necessary preconditions for the possibility of experience.[3] But Elkana plays down the importance of the relationship between Kant's philosophy and Helmholtz's scientific thought. The Kantian influence receives a mere four pages (pp. 165–9) while a full seventeen pages (pp. 71–88) are devoted to Humphrey Davy's poetic and scientific uses of the word 'energy'.[4]

5 Elkana tends systematically to underestimate the contributions of others to the discovery of the conservation law. This is particularly clear in his treatment of the origins of conservation in the analysis of the motive power of engines and concern with conversion processes.[5] As an example of the latter, Joule's

---

[1] *Op. cit.*, p. 117.

[2] The hypothesis of central forces is as Elkana rightly remarks 'the metaphysical core of the *Principia*' (p. 3).

[3] Kant [1775] and Kant [1786]. Heimann in his [1974*b*] gives a very detailed account of the relations between Helmholtz's metaphysics and Kantian philosophy, especially pp. 217–32.

[4] A few paragraphs would have sufficed to debunk the claim that Davy's (or Rumford's) works on the dynamical theory of heat were *necessary* preconditions for the discovery of the conservation law.

[5] Kuhn in his [1957] distinguishes three separate factors in the discovery of the law: the availability of conversion processes, concern with engines, and the philosophy of nature. Elkana criticises Kuhn's claim as to the simultaneous discovery of the conservation law (pp. 88–9 and again making exactly the same point on appendix pp. 178–9) on the grounds that different solutions were put forward to different problems. However, all the Kuhnian account claims is that from very different background problems, special cases of the general conservation law were discovered at around the same time. Elkana rightly makes the point that historians lack a general criterion for under what conditions a scientist may be said to have made a 'discovery' or discovered a law (p. 88 and again on appendix p. 178). But he then argues that no such general criterion exists

work on 'equivalents' gets particularly scant attention (pp. 115–6). After all Joule did write in 1847, as Elkana quotes, 'The general rule, then, is that whenever living force is *apparently* destroyed . . . an exact equivalent of heat is restored. The converse of this proposition is also true'. This seems to me to be a far more general and correct statement of the conservation law than Helmholtz's insistence on central forces. Elkana is quite right to point out that before Helmholtz's analysis conversion processes could not have been related in a general way to theoretical mechanics, in particular to the theorem of the conservation of *vis viva* (pp. 115–7).[1] Helmholtz first showed systematically how this could be done. Nevertheless it is clear that, particularly for Joule, Faraday[2] and Grove, the host of observed conversion processes and their concomitant equivalents formed the basis of their enunciations of the conservation law.[3]

6 Next I shall examine Elkana's criticism of what he designates as a 'standard' historical account of the discovery of the law of the conservation of energy. He remarks, 'One of the most cherished beliefs of many historians of science is that the principle of the conservation of energy grew directly out of the realisation of the impossibility of a perpetual motion machine' (p. 10). The principle of the exclusion of perpetual motion was enunciated in 1775 by the Acadèmie Royale de Science.[4] The problem then, as Elkana poses it, for this

---

and, 'for each specific case a criterion must be found'. But in what sense would one have a criterion, if it were unique to each specific case? (There seems to be little point in Elkana having reprinted his earlier paper ([1970*a*]) as an appendix to this volume since large portions of the main text are largely direct repetitions of sections of this earlier paper. For example pp. 88–91 of the main text repeat the pages 178–82 of the appendix.)

[1] As Elkana remarks (p. 90), Carnot realised that a problem was posed over the status of the principle of the exclusion of perpetual motion in areas outside mechanics. However Carnot regarded *all* action to be accompanied by motion, and thus perpetual motion could be universally excluded. He remarks, 'The objection may perhaps be raised here, that perpetual motion, demonstrated to be impossible by mechanical action alone, may possibly not be so if the power of heat or electricity be exerted; but it is impossible to conceive the phenomena to involve anything else but motion and as such subsist to the laws of mechanics'. (Carnot [1824], p. 12, n.*.)

[2] Elkana argues that Faraday employs the term force, in his principle of the conservation of force much in the same way as Helmholtz employs '*Kraft*' (pp. 130–2). Indeed this is correct if we concentrate *only* on the second sense of the conservation of 'force' in Helmholtz's paper, that of the indestructibility and interconvertibility of all agents which could affect an action. Faraday regarded force as a 'source of action', as is noted by Elkana (p. 131).

[3] Kuhn remarks of the significance of conversion processes in the discovery of the law that, 'In an important sense, . . ., the conservation of energy is nothing less than the theoretical counterpart of the laboratory conversion processes discovered during the first four decades of the nineteenth century'. (Kuhn [1957], p. 325). For an extensive analysis of conversion processes and their relationship to the discovery of conservation of energy, see especially Heimann [1974*a*].

[4] The text of the resolution is quoted by the author (p. 30). Of course the principle is very much older. (*Cf.* Mach [1883], chapter one and Mach [1872].) Mach argues that the major theoretical (or 'logical') root of the conservation law is the law of causality in the form that periodic or reversible changes in the elements of sensation cannot be induced by permanent changes therein. (Mach[1894], pp. 181–5). This law he claims is even older than mechanics. If this principle is taken as asserting that the idea of conservation is contained in the observation of reversible functional dependencies between the elements of sensation then it would appear to contradict the conservation

'standard' account is why did it take so very long (70 years) to formulate the full conservation law? Elkana's answer is that the impossibility of a perpetual motion machine is 'not a sufficient condition [for the principle of energy conservation] but only a necessary one' (p. 29). For energy might be *lost*, and thus not conserved, while the impossibility of a *perpetuum mobile* remained, *i.e.*, it would be impossible to *gain* motive power from nothing. Elkana actually provides us with an example of a working scientist who accepted both the impossibility of a *perpetuum mobile* and a *loss* of mechanical energy without compensation. Thus, 'William Thomson (later Lord Kelvin), while working on Carnot's theory and approaching the subject with the impossibility of perpetual motion in mind, at one and the same time lectured on electromagnetic induction, assuming without hesitation that in this case mechanical energy is simply lost without any compensation' (p. 10). Now while it is quite correct that the exclusion of perpetual motion mentions nothing of the interconvertibility of energy (this seems to me the crucial point) nor explicitly excludes its net loss, this conclusion does not contradict the assertion that the exclusion principle was a major precursor of the conservation law. Certainly the principle played a vital role in the analysis by, for example, Carnot and Clapeyron of the motive power of engines. It was systematically employed in the proof of a number of Carnot's major theorems, *e.g.*, that the efficiency of a Carnot cycle was maximal with respect to the class of cyclic transformations.[1] Carnot was explicit in the formulation he employed *viz.*: an 'exclusion of the possibility of an unlimited creation of motive power without consumption of caloric or of any other agent whatever'.[2] According to Kuhn,[3] Holtzmann and Colding as well as Helmholtz subsequently borrowed Carnot's technique of the analysis of cyclic processes employing the exclusion principle. Of the four scientists Kuhn mentions as being directly concerned with the motive power of engines, namely Hirn, Holtzmann, Seguin and Carnot, Elkana treats only the latter.[4] Indeed Elkana's own account of the role played by the exclusion of perpetual motion is unsatisfactory, for Helmholtz took the principle in the form of the impossibility of an 'unlimited

---

law by virtue of excluding irreversible processes at constant temperature. In the case of the mutual diffusion of two gases at constant temperature irreversible changes in the elements of sensation (corresponding to the diffusion) occur yet the total energy is conserved. (*Cf.* Mach [1894], p. 180).

[1] Carnot [1824], p. 12 and pp. 19–22 and Clapeyron [1834], p. 74. Clapeyron perceptively remarks that Carnot's discoveries were based upon the introduction into thermal physics of a new heuristic principle: 'The idea which serves as a basis of his researches seems to me to be both fertile and beyond question; his demonstrations are founded on the *absurdity of the possibility of creating motive power or heat out of nothing*. . . . This new method of demonstration seems to me worthy of the attention of theoreticians . . .'. (*Op. cit.*, p. 14).

[2] *Op. cit.*, pp. 11–12. There was a second major premise in Carnot's theory *viz.*: that the production of motive power involved the transportation of caloric, the re-establishment of caloric equilibrium and not its consumption. Both assumptions play crucial roles, removing *either* would destroy the validity of Carnot's arguments. *Cf.* Elkana's comments, p. 88–91.

[3] Kuhn [1957], pp. 334–35.

[4] Elkana claims that the thesis that conservation ideas grew out of a concern with engines, is dealt with in his chapter on 'Heat and Energy' (p. 114). But I can find no mention of the subject there apart from his remarks on Carnot and the *Second* law of thermodynamics.

creation of motive power' as *both* a necessary and sufficient condition for his theorem on the conservation of energy. As has already been noted (p. 166 *above*) Helmholtz is explicit in his use of the exclusion principle as a premise. Commenting on this point Elkana states that, 'For Helmholtz the true meaning of this impossibility was that any conservation law must imply the impossibility of perpetual motion' (p. 130); and later 'As Helmholtz formulates it, he did not look for a law which would be a generalisation of this impossibility. He knew only that whatever law he was going to suggest must not contradict this impossibility' (p. 159). But surely this is quite wrong, given Helmholtz's *explicit* statements about the assumption he employs and the logical form of his argument. Elkana remarks of this problem: 'Do we have to take Helmholtz at his word, accepting that the whole idea had its basis in an inductive realisation that one could not construct a perpetuum mobile?' (p. 159). But of course we can take Helmholtz at his word without having to accept that any 'inductive realisation' was involved. Elkana answers his question in the negative, but the problem here is not where did the idea of conservation originate, but on what did Helmholtz base his proof. What relationship did he claim existed between the principle of exclusion of perpetual motion *and* the general conservation of 'force' law? Indeed as we have noted Helmholtz claims that the conservation of 'force' law is equivalent to the universality of central forces. Further, if some forces depend upon the velocity Helmholtz remarks that 'combinations of such bodies would be possible in which force might either be lost or gained *ad infinitum*'. Hence the negation of the central force hypothesis implies the negation of the exclusion principle. But the central force hypothesis is equivalent to the conservation of 'force', whence the negation of the conservation of 'force' must imply the negation of exclusion of perpetual motion. Whence on Helmholtz's account the exclusion principle must imply the conservation of 'force'.[1]

7 Finally, let me turn to the general concept 'image of science' with which Elkana hopes to improve our historical understanding. Elkana locates the fundamental historical problems as, 'what causes change in the contents of knowledge?' and 'what is that part which "grows"?' (p. 6).[2] His own approach to these problems relies upon in each case under consideration identifying and analysing the interaction among three distinct factors. They are (pp. 7–8); (*i*) Developments in the body of knowledge, (*ii*) 'The social image of science in a given place or community at a given time: whatever people in general and scientists in particular *think* consciously of science, its role, its ethos *etc.*'; (*iii*) Social and political factors. The extent to which this guide to the writing of history differs from the more traditional analysis into 'internal' and 'external'

---

[1] Elkana gives a second argument for his claim (p. 136). 'The way I understand Helmholtz's approach to the principle of the impossibility of a perpetual motion machine *i.e.* as a necessary implication of the basic conservation law—once it is proven, is that nothing is more natural than to treat this principle as one of "those fundamental theorems" the exact status of which has not yet been clearly established in mechanics?' But proven from what principles? The author remarks 'I did not refer to the impossibility of perpetual motion as one of Helmholtz's premises, since I do not think that it occupied an important place among his presuppositions, in spite of what he says . . .' (p. 130). But the structure of Helmholtz's *argument* belies this comment.

[2] One would have thought that the two questions could be answered simply by replying to the first 'inconsistencies' and to the second 'systems of theories'.

history coincides with the stress the author places on what scientists think consciously of science, *i.e.* (*ii*) *above*, yet in the writing of his case-study Elkana often remarks that scientists' comments on their work and its methodological significance need not be taken at face value. As has been noted Elkana chooses to ignore Helmholtz's claims as to exactly on what he bases his proofs (*above*, p. 174), he describes Carnot as an example of a scientist who 'pays lip-service to a methodology which partially or even completely belies [his] own conceptual framework' (p. 90).[1] The author even quotes Einstein to the effect that one shouldn't listen to scientists' words but fix attention on their deeds. But then what particular value can their 'image of science', their methodological appraisals, have for the writing of better history of science? Has the 'image of science' improved the historical accounts of the discovery of the conservation of energy? I should say not, for it has led the author into making unsupportable claims. For example after an interesting analysis of the institutional setting in which work on the problems of the conservation of energy was undertaken in France, England and Germany and the diverse prevailing attitudes as to exactly what constituted 'good' and 'bad' science, Elkana makes the claim that 'The law of conservation of energy *could not* be born either within the institutional framework of France or that of England' (p. 155, my italics). One wonders what possible empirical evidence could the author bring to support such a claim? It is anyway plainly false, given for example the work of Joule or Faraday in England.

Elkana continues of the institutional setting in Germany, 'What I do claim is that exactly the atmosphere which was detrimental to many other scientific problems . . . was indispensable for the establishment of the conservation of energy principle *i.e.* the emergence of the concept of energy' (p. 155–6). But if this was really so, why was Helmholtz's paper rejected by the *Annalen der Physik* and why was there so much reservation in Germany in accepting the significance of Helmholtz's work,[2] if the situation was so ripe for the establishment of the law?

This book fails in its main aims. Despite the fact that Elkana makes a number of interesting asides there is little in the case study which adds to our historical understanding of the discovery of the law of the conservation of energy. Indeed the author's philosophical preconceptions have in my view lead him astray on a number of crucial historical points. This in turn has produced an incorrect appraisal both of Helmholtz's achievement and its significance for the theoretical physics of the time.

PETER CLARK
London School of Economics

REFERENCES

CARNOT, S. [1824]: 'Reflections on the Motive Power of Fire and on Machines Fitted to Develop that Power', reprinted in E. Mendoza (*ed.*): *Reflections on the Motive Power of Fire*, pp. 3–59.

CLAPEYRON, E. [1834]: 'Memoir on the Motive Power of Heat', reprinted in E. Mendoza (*ed.*): *Reflections on the Motive Power of Fire*, pp. 73–105.

---

[1] This claim is not argued. What does Elkana take to be Carnot's methodology? Where does Carnot pay this lip-service to it?

[2] An account of the immediate reception of Helmholtz's paper is given in Koenigsberger [1902], p. 37–51.

CLAUSIUS, R. [1853]: 'Über einige Stellen der Schrift von Helmholtz "Über die Erhaltung der Kraft"', *Annalen der Physik und Chemie*, **89**, pp. 574–7.

CLAUSIUS, R. [1854]: 'On a Modified Form of the Second Fundamental Theorem in the Mechanical Theory of Heat', reprinted in T. Archer Hirst (*ed.*): *The Mechanical Theory of Heat*, pp. 111–35.

DORLING, J. [1970]: 'Maxwell's Attempts to Arrive at Non-Speculative Foundations for the Kinetic Theory', *Studies in History and Philosophy of Science*, **1**, pp. 229–48.

ELKANA, Y. [1970*a*]: 'The Conservation of Energy: A Case of Simultaneous Discovery?' *Archives Internationales d'Histoire de Sciences*, **24**, pp. 31–60.

ELKANA, Y. [1970*b*]: 'Helmholtz's "Kraft": An Illustration of Concepts in Flux', *Historical Studies in the Physical Sciences*, **2**, pp. 263–98.

ELKANA, Y. [1971]: 'The Problem of Knowledge', *Proceedings of the Second Greek Humanistic Symposium*, pp. 191–247.

ELKANA, Y. [1975]: 'Boltzmann's Scientific Research Programme and its Alternatives', in Y. Elkana (*ed.*): *The Interaction Between Science and Philosophy*, pp. 243–79.

HEIMANN, P. M. [1974*a*]: 'Conversion of Forces and the Conservation of Energy', *Centaurus*, **18**, pp. 147–61.

HEIMANN, P. M. [1974*b*]: 'Helmholtz and Kant: The Metaphysical Foundations of *Über die Erhaltung der Kraft*', *Studies in History and Philosophy of Science*, **5**, pp. 205–38.

HELMHOLTZ, H. VON [1847]: 'On the Conservation of Force; a Physical Memoir', reprinted in J. Tyndall and W. Francis (*eds.*): *Scientific Memoirs, Natural Philosophy*, pp. 114–62.

KANT, I. [1755]: *Allgemeine Naturgeschichte und Theorie des Himmels.*

KANT, I. [1786]: *Die metaphysischen Anfangsgründe der Naturwissenschaft.*

KOENIGSBERGER, L. [1902]: *Hermann von Helmholtz.*

KUHN, T. S. [1957]: 'Energy Conservation as an Example of Simultaneous Discovery', in M. Clagett (*ed.*): *Critical Problems in the History of Science*, pp. 321–56.

LIEBIG, J. [1842]: *Thier-Chemie, oder die Chemie in ihrer Anwendung auf Physiologie und Pathologie.*

MACH, E. [1872]: *Über die Erhaltung der Arbeit.*

MACH, E. [1883]: *Die Mechanik in Ihrer Entwicklung Historisch-Kritisch Dargestellt.*

MACH, E. [1894]: 'On the Principle of the Conservation of Energy', reprinted in *Popular Scientific Lectures*, pp. 136–85.

MAXWELL, J. C. [1865]: 'A Dynamical Theory of the Electromagnetic Field', *The Scientific Papers of James Clerk Maxwell*, **1**, pp. 526–97.

WEBER, W. [1869]: 'Über einen einfachen Auspruch des allgemeinen Grundgesetzes der elektrischen Wirkung', *Wilhelm Webers Werke*, **4**, pp. 243–6.

WEBER, W. [1878]: 'Elektrodynamische Maasbestimmungen insbesondere über die Energie der Wechselwirkung', *Wilhelm Webers Werke*, **4**, pp. 361–419.

DREYFUS'S DISPROOFS*

Dreyfus's book[1] is based on a memorandum[2] of some years earlier. His thoughts on Artificial Intelligence at that time received some philosophical attention from those few philosophers who were prepared to master the details of the empirical research Dreyfus described and criticised. But from the Artificial Intelligence (*AI*) workers themselves Dreyfus provoked a great deal of anger but very little published reply, largely because few were willing or able to reply to Dreyfus's philosophical arguments in their own terms, and the fact that those terms were drawn from continental phenomenology did not help matters.

The heart of Dreyfus's case is that *AI* work is not only unsuccessful in fact in its aim of simulating intelligent behaviour on digital computers, but that this task cannot in principle be carried out. In reply, *AI* workers have contented themselves with what one might call Johnsonian refutations—remembering his treatment of Berkeley's views on the non-existence of material substance by kicking a stone in the street. Dreyfus played a mechanical chess player in 1969, was beaten, and for many *AI* people that settled the matter. In the present book Dreyfus describes the background to those events a little ungraciously—he had an initial encounter which he won, and then later, 'Embarrassed by my exposé . . . *AI* workers finally produced a reasonably competent program' (p. xxxii). He recounts how that program reached a peak of performance after which it was beaten a lot. This, he adds, is just what one would expect. What he does not say is that, given his presupposition that there cannot be intelligent machines, their coming up with that 'reasonably competent program' is exactly what one would not expect.

Dreyfus has three general lines of argument supporting his claim:

(1) An examination and criticism of actual *AI* systems devised and programmed between 1957 and 1967.
(2) A discussion of four assumptions, the Biological, the Psychological, the Epistemological and the Ontological respectively, that Dreyfus claims underlie all *AI* work.
(3) General arguments, on phenomenological grounds, against the very possibility of the *AI* enterprise.

There is a combination of empirical and *a priori* arguments here familiar since Descartes argued the impossibility of a vacuum in nature and pointed out that there was in fact always air found in the Torricellian tubes.

I shall argue that Dreyfus has not in fact described the *AI* enterprise correctly. Secondly, that actual work in *AI* has changed considerably since his survey of it stopped, and that it is beginning to incorporate many of the features that he claimed an adequate *AI* system must embody, though none did at the time he wrote. Thirdly, that there is a lack of clarity about exactly what Dreyfus is claiming, because the word 'digital' is an essential part of his impossibility claim, in that he is not claiming that intelligent behaviour could not be simulated by any machine whatever, and indeed he hints that it might be possible to simulate it with analogue machines. Hence, much turns on the difference between digital and analogue machines, and I am not at all sure that Dreyfus has got it right. Lastly, I shall argue that Dreyfus's one clear *a priori*, phenomenological, argument proves, on examination, far more than he would wish. I will make these points in the course of discussing Dreyfus's three forms of argument in turn, and as listed and numbered above.

---

* Review of H. Dreyfus [1972]: *What Computers Can't Do*. New York: Harper, pp. 259.

[1] Dreyfus [1972].

[2] Dreyfus [1965].

Of the first, Dreyfus's critical discussion of early *AI* work, it must be said that it has had a very salutary effect, and should be required reading for everyone in the field he criticises. The naive, and from this distance in time, absurd over-optimism of many in early *AI* is now sad reading, and Dreyfus is quite right to expose it. Where Dreyfus goes too far is in moving from the naivety of very clever men in the face of enormous and unrecognised problems to accusations (*ibid.*, p. 223) of actual and conscious dishonesty.

Dreyfus enlarges his target to cover early work in machine translation (*MT*) and here the first of a number of curious failures of judgment on the author's part becomes evident. It is true that the 'sociology' of the *MT* work had much in common with that of the later *AI* work, and Dreyfus makes much of the similarity: early over-confidence followed by later disillusion, though the latter was deeper and more lasting in the *MT* field.

But the difference between the two enterprises is what is important, since Dreyfus's book is about *AI*, and their intellectual presuppositions are more important than the psychological histories of their respective researchers. From the very beginning, *AI* workers emphasised the importance of interactive, self-modifying mechanisms in research on the structure of mental activity, and of the importance of an adequate manipulation of knowledge about the real world. The early *MT* work had no place for either of these and their absence was one solid reason why the *MT* enterprise failed and failed clearly. We simply cannot translate complex sentences which, say, refer pronouns unless we can manipulate knowledge of what is the case. Anyone who doubts this should think about how he knows the correct referent of 'several' in 'The soldiers fired at the women and I saw several fall'.

Such features are essential to *AI* work and so no criticism by Dreyfus of early *MT* work is to the point at all. No one in *MT* in the Fifties considered himself engaged in the simulation of 'artificial reason' and Dreyfus might as well have included within his criticisms any other field, like control engineering or operations research, in which the computer plays a large part.

The same sort of large-scale misjudgment appears when Dreyfus claims that the success of Chomsky's linguistic enterprise (*ibid.*, p. 110) '. . . is a major source of encouragement for those in AI who are committed to the view that human behaviour can be formalised without reduction to a physical level'.

There are two odd things there. In the first place, Chomsky's work has had little influence on *AI* work. On the contrary, there was an implicit agreement among *AI* people almost from the beginning that work that had no place for any notion of meaning or facts, and whose view of language structure was that each and every sentence was derived in isolation from a discourse or other use, could be at best a poor guide to the nature of human communication and understanding. (There was some brief attempt to use Chomskyan notions to explore the nature of vision, but they came to grief for the same reasons as in the language case, albeit later.) One of the motivations of *AI* work, at the time Dreyfus describes and since, has been to show that one could do better than Chomsky in structuring natural language without his naive and rigid theoretical framework.

A second oddity is the general attitude to Chomsky's work in Dreyfus's book. It is described (*ibid.*, p. 110) as a 'breakthrough' in that 'A segment of orderly behaviour which at first seems non-rulelike turns out to be describable in terms of complex rules, rules of the sort that can be processed directly by a digital computer'.

Yet, from all Dreyfus says against the possibility of the simulation of such sensitive areas of human behaviour by systems of rules embodied in a computer, one would expect him to subject Chomsky to the same sort of 'phenomenological criticism' as *AI*: to argue that Chomsky's proposals did no sort of justice to the objective, phenomenological, reality of our use and understanding of language 'from inside'.

Yet, he does not. Indeed, Dreyfus subjects Chomsky, and in particular his 'competence-performance' distinction,[1] to less fundamental criticism than any philosopher who has ever had cause to discuss Chomsky's work.

This curious judgment is explained perhaps by the fact that Chomsky has never claimed that his bodies of rules, if actually constructed, would model, represent, or simulate any human activity whatever. The 'competence-performance' distinction was in fact created by Chomsky to define an area of 'competence', expressed by the linguistic rules, but without making any claims whatever about actual human performance, and therefore not connected, of course, with any possible tests or observations. Thus, Dreyfus is accepting Chomsky's work because it makes no such claims, while condemning *AI* work because it does: (p. 110)

> Linguistic competence is not what *AI* workers wish to formalise. If machines are to communicate in natural language . . . they must . . . contain rules of linguistic performance . . . (but) is there reason to think there can be a formal theory of performance? There are two sorts of reasons to think such a . . . theory is impossible . . . (2) a descriptive objection (to which we shall now turn): not all linguistic behaviour is rulelike. We recognize some linguistic expressions as odd—as breaking the rules—and yet we are able to understand them.

There are two confusions here and it is important to sort them out. First, there is the limited notion of 'formalisation' apparently in Dreyfus's mind. By that he seems to understand rules of general validity, without exceptions, and the later context suggests he means without meta-rules, ranging over object rules, either.

But this is too limited by far; a rule with a finite list of exceptions is no less formalised! If Dreyfus means by 'formalised' simply what can be put in a consistent manner into a computer program, a pretty good definition in this area, then he is straightforwardly wrong.

Secondly, and more importantly, is Dreyfus's uncritical acceptance of Chomsky's competence-performance distinction in order to make points against *AI*. This is no place for more critical discussion of Chomsky's distinction, but if one accepts the following then nothing along the lines Dreyfus wants can possibly follow. Competencies, whatever they may be, can be nothing above and beyond selections from among possible performances, for they must attach somewhere to assessable realms of behaviour.[1] The generative semanticists[2] have perhaps belatedly recognised this by claiming (for their version of Chomsky's theories) that the relations they postulate between sentences exist in an abstract realm of relations like those of logic. This is again not the place to discuss such rescue plans for Chomskyans, but let us look here and now at the example Dreyfus gives in the quotation. In the ordinary sense of 'competence' he is, in drawing attention to our ability to understand 'rule breaking utterances', remarking on a human competence not explicated or covered by Chomsky's theory of competence. But no new realm of performance has been entered—one competence has merely been contrasted with others. Once we see that, the whole of Dreyfus's case against *AI* based on comparison with Chomsky's work simply falls away.

---

[1] Chomsky [1965].

The point can be made clear by seeing how one prominent *AI* worker, John McCarthy,[3] defines the task of *AI* as: 'studying the structure of information and the structure of problem solving processes independently of applications and independently of its realization in animals and humans'.

Two things emerge from that quotation: first that Dreyfus misrepresented at least some principal *AI* workers when he defined *AI* only as the simulation of various forms of behaviour: and secondly, and to the present point, McCarthy's definition of *AI* is highly reminiscent of Chomskyan competence, and makes it hard, if not impossible, for Dreyfus to contrast the two in the way he wishes to.

The two positions, Chomsky and *AI*, are quite different in other ways of course: not only in the far greater flexibility and sophistication of the methods of *AI*, but chiefly because it retains the testable contact with reality that Chomsky has denied his own system by the construction of the competence-performance distinction. For it is an unspoken assumption of McCarthy, like all *AI* workers, that the rules constructed should be implemented on a machine to see if they produce the desired result. This does indeed produce some machine performances, but no firm claims are made about the theoretical status of such performances, nor need any be made. The contrast is clear with Chomskyan linguistics, where any such implementation, or test of even the consistency of such bodies of rules, would move one into the wholly other and wholly bad realm of 'performance'.

I shall now move to the second major area of Dreyfus's discussion: the four assumptions 'underlying current *AI* work', and argue in the same vein that he misrepresents the work that actually goes on. The four are:

(1) THE BIOLOGICAL ASSUMPTION: '... the naive assumption that man is a walking example of a successful digital computer program'.

(2) THE PSYCHOLOGICAL ASSUMPTION: that '... the mind functions like a digital computer'.

(3) THE EPISTEMOLOGICAL ASSUMPTION: that '... although human performance may not be explainable by supposing that people are actually

---

[1] This point is discussed in detail in Wilks [1974].

[2] See for example Lakoff [1971].

[3] McCarthy [1974].

following heuristic rules in a sequence of unconscious operations, intelligent behaviour may still be formalizable in terms of such rules and thus reproduced by machine'.

(4) THE ONTOLOGICAL ASSUMPTION: that '. . . everything essential to intelligent behaviour must in principle be understandable in terms of a set of determinate independent elements . . .'.

We can ignore Dreyfus's first two assumptions if we accept McCarthy's view as a correct report on at least his own views, though it is in fact of wider value, since *AI* workers simply do not in general make those assumptions, nor, as Dreyfus repeatedly claims, need they do so in order to do *AI*. As to the second pair, before we can see if *AI* workers hold them, we must first clarify them a little to see what, if anything, is being held. The fourth is particularly obscure. Couched as it is in terms of the notion of understanding, it would seem to be an epistemological assumption, in the normal philosophical meaning of that term as having to do with knowledge and the knowing process. Again, it is hard to know what it would be like not to make the assumption in some form or other, for, as Margaret Boden has pointed out[1] a propos of Dreyfus's argument 'Dreyfus himself uses the twenty-six letters of the alphabet to represent his "ambiguous and indeterminate" information.' There is undoubtedly a version of the fourth assumption that is held by some, though not all, AI workers, and equivalent to "the world is an assemblage of facts".' Notice that this assumption, unlike Dreyfus's, really is ontological, but it is by no means a required belief for workers in *AI*, and indeed belief in it is probably waning.

As to the third 'epistemological' assumption, it again seems to suffer from naming trouble, since it is in fact no more than a methodological postulate. It is not an assumption about any part of the real world, but about the nature of the *AI* task. And Dreyfus may well be right here that many if not all *AI* workers do accept it. But I would claim that Dreyfus nowhere shows in his book that they are mistaken. He shows the short-comings of some early *AI* work, but in the six years that have passed since his survey came to an end, things have improved in many areas. Indeed, and this is part of the comedy of the debate, Dreyfus actually lists aspects of human information processing that he says would have to be taken account of in any successful *AI* enterprise: fringe consciousness, ambiguity tolerance, essential/inessential discrimination, and perspicuous grouping. He also adds, in his detailed discussion, the notion of the perception of *gestalten* or wholes. But, if he were to resurvey the field now, he would find the beginnings of the implementations of some of these very features, particularly the notion of pattern seeking wholes in the analysis of natural language.[2] The work I refer to is all very primitive at the moment, but the importance here is that Dreyfus could not simply repeat his empirical criticisms were he to resurvey the field, and that alone is considerable evidence that he has not demonstrated that any task could not be done, but only that it has not been done. Moreover, the assumptions he characterised as essential to *AI* were in no case so, and the only one that was widely and generally held, the 'epistemological assumption' has not been shown to be fundamentally wrong.

What empirical matters, then, are at issue between Dreyfus and those whom he attacks? Very few in fact, once we see that the field is changing, and that the assumptions Dreyfus considers essential to the *AI* task are not in fact so. Apart from Dreyfus's *a priori* arguments to which I shall come in a moment, the only empirical issue concerns the difference between analogue and digital computers. It is important to see, what a casual reader might miss, that Dreyfus at no point claims that machines cannot simulate intelligent behaviour, only that digital computers cannot. And, in my view, he produces no empirical arguments that they cannot, only that they have not. But what then is it that causes him to believe that analogue computers might do better, and is he not contradicting himself, given that there is a proof that any analogue computer can be reproduced, as regards its input and output, to any desired degree of closeness, by a digital one?

---

[1] Boden [1972], p. 148.

[2] See, for example, Charniak [1975], Schank [1975], Wilks [1975].

The first of these questions has been discussed in some detail by Pylyshyn and Boden. As to the second, Dreyfus's arguments on the matter are odd, to say the least. He produces a complex argument that although a digital computer may 'simulate' the behaviour of an analogue device like a slide rule, it cannot 'represent' or 'reproduce' it because the two are processing 'different information'. The reader must judge for himself the value of such arguments, but he should be wary of accepting them from one who, in another context, has attacked others for using the notion of information in an unquantified informal way (*ibid.*, p. 79).

Pylyshyn[1] has pointed out that Dreyfus is not accepting normal canons of what it is for one process to explain or reproduce another, even though (1) Dreyfus gives no description of what it would be for one process to reproduce another without being identical to it, (2) he explicitly disclaims that he is demanding identity between explanans and explanandum (*ibid.*, p. 236), and (3) part of what it is for one process or entity to explain another is that they are not identical, for if they are nothing is explained.[2]

But there may be a loophole in this anti-Dreyfus argument: for like all such formal proofs, the result in question—the equivalence of digital and analogue processes—holds only for the class of objects defined in the proof. There may well be entities or systems that are analogue devices in the commonsense meaning of the term, though not covered by the result. In particular, it is still an open question whether or not the three-body problem can even be recursively approximated by digital means. If it should turn out not to be so, then any three bodies, in virtue of their mutual gravitational attraction, would form a device not simulable by a digital computer. It may be a little hard to see how such a device could be used to simulate human behaviour, but it would constitute an important theoretical loophole in that particular argument against Dreyfus's general claims.

The latter part of the book assembles a number of quotations and arguments of a quite different sort. Merleau-Ponty, Heidegger, Polanyi and Wittgenstein are among those called in to tie off the arteries of rationalism, to argue, as the phenomenologist wishes to, that, contrary to what scientists tell us, the world is much as it seems, one of people and feelings and meanings and understandings.

A little translation can make such arguments seem cosy and familiar to anyone brought up on ordinary language philosophy or Moorean common-sense analysis. There are not two tables, the atomic and the everyday, but only the one I normally eat off. The Wittgensteinian form-of-life clinches the matter: the world really is how our life shows it to be, and our words mean what they normally mean which is what they are learned in normal circumstances as meaning, hence a computer without a life-form, a body or an upbringing cannot participate in that world and cannot be intelligent.

---

[1] Pylyshyn [1974].

[2] A point explored in Boden [1961].

There is much of value in Dreyfus's arguments and it is useful to have them restated against the well-known counter claims of computer scientists. But I am not at all sure that these considerations will or can provide the *a priori* proof of the impossibility of machine intelligence that Dreyfus so wants.

Let us take one very simple consideration about the notion of understanding, say. For Dreyfus my understanding of things is rooted in my own moments of understanding—I know what it is to understand because I do sometimes understand. The phenomenological argument, about how things appear to me, is at one with the ordinary language argument at that point.

As Dreyfus puts it when discussing Bobrow's[1] STUDENT program (*ibid.*, p. 46–7):

> If we remember that 'understands' merely means 'answers questions in a restricted subset of English subject to algebraic constraints', then we will also remember that although the words in quotation marks have nothing to do with what human understanding normally means, they are nonetheless accurate.... Minsky exploits this ambiguity in his rhetorical article by simply dropping the quotation marks.

But the real trouble is still there tucked away in our 'form of life', for this feeling for how things appear may justify my own understanding, but how does it justify anyone else's? If I am asked about the phenomenology of anyone else's understanding, I have, of course, no feelings and immediate assurance to fall back on, and I am out in the cold world of watching his behaviour for appearances of understanding, and searching his expressions in the approved manner for those who worry about 'other minds'.

The point should now be clear, and in Dreyfus's own terms: what kind of phenomenological reassurance about the understanding of another can I have that I do not have for a machine—the criteria are and must be behavioural in both cases. The only escape is by an argument from structural analogy: that the other person is warm-blooded and brain possessing, like me and unlike the apparently intelligent machine. But, of course, the phenomenologist is the last person who can fall back on a crude argument from physiological analogy, one that does not provide the sort of evidence for your understanding that I have for mine. And that alas we know we cannot have.

The phenomenologist may argue at this point that his knowledge of the understanding that other persons have rests, not on any direct experience by him of their minds, but upon his own experience of, say, conversation with them. But to that I reply that, as a matter of fact, such an experience does not distinguish other persons from machines. There are already 'conversational programs' in existence[2] that leave many uncertain, after some written conversation with them by means of mechanical typewriters, whether or not they are conversing with a person.

The phenomenologist, who argues on the basis of how things appear to be, is now in a difficult position, and may be forced to say that what seems to be a conversation cannot really be one if an entity participating in its turns out to be a computer, all of whose operations reduce in fact to manipulations of electronic 'bits'. But this once again is a wholly unphenomenological argument appealing beyond appearance to internal structure.

---

[1] Bobrow [1968].

[2] Colby [1971].

A more sophisticated reply from the phenomenologist would be to the effect that a person participating in such a conversation with a computer is simply being fooled, and there is nothing that prevents a phenomenologist from being misled, just as one may think one hears live music in a further room, which, on entering, turns out to be a record playing.

But if we look harder at what one so misled *is being misled about,* we see that the phenomenologist does not succeed in establishing the distinction he needs, because the person who finds a record player is indeed being misled in the same way as one who thinks he is talking to a person but is really talking to a machine, but they are not misled, respectively, about the hearing of music or the having of a conversation. And it was on the experience of having a conversation that the experience of the understanding of others was originally said to rest.

My final point then is that Dreyfus's empirical arguments are not sound, they disprove nothing, while his one solid *a priori* argument just above proves far too much. It shows that machines cannot understand, but at the old philosophical price of showing that no one but I can either.

Whether or not its survey of research is fair, or its arguments are sound, the book has done a great deal of good in a field whose lack of self-generated criticism is scandalous. Dreyfus should review the field every five years and change his empirical arguments each time—*AI* workers should then be grateful and ask no more of him, for he would be doing them a considerable service.

The point in the field where Dreyfus's attention is needed at the moment is that part of Artificial Intelligence research directed towards the understanding of natural language, some of which is, in my view, making use of rules of inference that express a notion very close to what Wittgenstein meant by a form of life, and which are hinted at in passages like[1]:

> And if things were quite different from what they actually are—if there were for instance no characteristic expression of pain, of fear, of joy; if rule became exception and exception rule; or if both became phenomena of roughly equal frequency—this would make our normal language games lose their point.

It seems to me, and this cannot be argued in detail here, that Wittgenstein is drawing attention to a vital aspect of our understanding of natural language, and one which has been totally ignored by the generative linguists. The simplest of examples will show the need. Let us consider 'John saw the figure in the cloak coming for him with a knife, and he swung his body to the right to avoid it'. A reader knows quite well that the last word in the sentence refers to the knife, and not to any of the other three objects mentioned. And yet, it is not possible to specify how that reference is made (one which might be vital for translation of the sentence into another language in which those nouns had different genders) without knowing not only detailed facts about the objects mentioned, but also very general 'facts' about our normal form of life, such as that people do not want to get hurt. If so universal a fact were to change, our understanding of language would change with it. There is a current dispute among artificial intelligence workers as to whether such knowledge is always detailed knowledge of areas of discourse[1] or whether there is some central core[2] of such assumptions required for the understanding of language.

---

[1] Wittgenstein [1953], I, para. 142.

Here then is a contribution to the understanding of language, coming from *AI* and not from linguists or logicians, which Dreyfus ought in my view to welcome as contributing to the sort of study of language he would like to see undertaken.

YORICK WILKS
Istitut per gli Studi Semantici e Cognitivi,
Switzerland

REFERENCES

BOBROW, D. [1968]: 'Natural Language Input for a Computer Problem-Solving System', in M. Minsky (*ed.*): *Semantic Information Processing.*

BODEN, M. [1961]: 'The Paradox of Explanation', *Aristotelian Society Proceedings, 1961–62*, pp. 212–29.

BODEN, M. [1972]: *Purposive Explanation in Psychology.*

CHARNIAK, E. [1975]: 'A Partial Taxonomy of Knowledge about Actions', in (*ed.*): *Proceedings of the Fourth International Conference on Artificial Intelligence.*

CHOMSKY, N. [1965]: *Aspects of the Theory of Syntax.*

COLBY, K. *et al.* [1971]: 'Artificial Paranoia', *Artificial Intelligence*, **2**, pp. 1–25.

DREYFUS, H. [1965]: Alchemy and Artificial Intelligence, Rand-Memorandum, P–3244.

DREYFUS, H. [1972]: *What Computers Can't Do.*

LAKOFF, G. [1971]: 'Generative Semantics', in D. Steinberg and L. Jakobovits (*eds.*): *Semantics.*

McCARTHY, J. [1974]: Review in *Artificial Intelligence*, **5**, pp. 317–22.

PYLYSHYN, Z. [1974]: 'Minds, Machines and Phenomenology', in *Cognition*, **3**, pp. 20–42.

SCHANK, R. (*ed.*) [1975]: *Conceptual Information Processing.*

WILKS, Y. [1974]: 'One Small Head', in *Foundations of Language*, **11**, pp. 77–95.

WILKS, Y. [1975]: 'An Intelligent Analyser and Understander of English' in *Communications of the A.C.M.*, **18**, pp. 264–74.

WITTGENSTEIN, L. [1953]: *Philosophical Investigations.*

---

[1] As is assumed by Charniak [1975].

[2] As is assumed by Schank [1975] and Wilks [1975].

Brit. J. Phil. Sci. 27 (1975), 187–197 *Printed in Great Britain*

Reviews

DRAKE, F. R. [1974]: *Set Theory: An Introduction to Large Cardinals*. Amsterdam: North-Holland. $23.25. Pp. xii+350.

JECH, T. J. [1973]: *The Axiom of Choice*. Amsterdam: North-Holland. $15.50. Pp. ix+202.

These works contain systematic accounts of important research in set theory carried out mainly in the 1960s. Jech's book is devoted to applications of Cohen's forcing method to the problem of independence of various forms of the axiom of choice, while Drake's provides an account of the 'strong axioms of infinity' and their consequences. Both works are the first accounts of their respective subjects in book form.

The *axiom of choice* (*AC*)—which asserts that, for each non-empty family S of non-empty sets, there is a function (called a *choice function* on S) which picks out a member of each set in S—was first explicitly formulated by Zermelo in 1904 to show that every set can be well-ordered. The axiom met with a certain amount of scepticism from mathematicians, notably the members of the French semi-intuitionist school centred round Borel. The chief objections to the axiom were based on the following grounds: (1) its 'non-constructive' and purely existential character: it does not indicate how the choice function in question is to be obtained, or indeed characterise it in any other way; (2) its 'implausible' consequences, *e.g.* Hausdorff's discovery that it implies that 1/3 of a sphere can be congruent to 2/3 of it. It did not take very long, however, for most mathematicians to realise that it was destined to play an essential role in the increasingly abstract mathematics of the twentieth century. Indeed, it turned out that many important theorems whose proofs depended on the axiom were actually *equivalent* to it, so that rejecting the axiom meant rejecting the theorems, a course which few mathematicians were willing to countenance, let alone take. Thus it seems fair to say that the axiom achieved respectability largely on *pragmatic* grounds (although a number of mathematicians, for example Zermelo himself, had been prepared to accept it *ab initio* in virtue of its 'self-evident' character!). Formal confirmation of its respectability had to wait, however, until 1938 when Gödel established its *relative consistency* with the other axioms of set theory, *i.e.*, if these latter do not lead to contradiction, then AC can be safely added without engendering any new contradictions. The question of its *independence*, although a moral certainty owing to the pioneering work of Fraenkel and Mostowski, remained formally open until 1963 when Cohen invented his celebrated method of forcing and used it to show that the axiom of choice and several other central principles and conjectures of set theory (notably the continuum hypothesis) cannot be proved from the other axioms, assuming the consistency of these latter.

I pointed out earlier that many results in mathematics whose proofs require the axiom of choice turn out to be *equivalent* to it. This is not always the case, however. For example, consider the Gödel–Henkin completeness theorem for first-order logic in the form: every consistent set of first-order sentences (of arbitrary cardinality) has a model. Any proof of this theorem uses the axiom of choice, so we may say that the axiom of choice (in the presence of the other axioms of set theory) *implies* the completeness theorem. But it is known that the completeness theorem, although not *itself* provable from the other axioms of set theory, does *not* imply the axiom of choice. In general, let us call a set-theoretic assertion A a *weak version* of the axiom of choice if (1) the axiom of choice implies A, (2) A does not imply the axiom of choice, and (3) A is not provable from the remaining axioms of set theory. Then we may say that the completeness theorem is a weak version of the axiom of choice. Many other weak versions of the axiom of choice are known, for example, the existence of a non-Lebesgue measurable set of real numbers, and the axiom of choice for families of finite sets.

Jech's book provides a comprehensive account of the work I have outlined above. The discussion is technical and somewhat compressed, and is mostly intended for mathematical specialists. Chapters 1, 2 and 10, which are devoted to applications of the axiom of choice in mathematics and to describing the curious situation that arises when the axiom is dropped, are of a more general character than the others and could be usefully studied by anyone modestly acquainted with naive set theory and real analysis. In the remaining chapters the author discusses the consistency and independence of the axiom (using the methods of Gödel and Cohen and their extensions); weak versions of the axiom; the relation between Cohen's independence technique and the earlier one of Fraenkel and Mostowski (which established the axiom's independence of a system of set theory weaker than the one customarily used); the behaviour of cardinal numbers without the axiom; and some assertions contradicting the axiom. The book contains lots of (mostly difficult) exercises. To sum up, this is a very useful book for specialists.

Broadly speaking, the results in set theory obtained since the early 1960s can be assigned to two main categories. The first consists of independence results obtained by means of Cohen's forcing technique, and the second is the subject of large cardinals, which was chiefly created by Tarski and his students in Berkeley. Now there is no hard and fast definition of 'largeness' for cardinals, although many cardinal-theoretic properties are known to confer some attribute of 'largeness' on their possessors. In order to clarify this it seems best to give an example.

There are two basic methods built into the axioms of set theory for obtaining new, bigger, cardinals from smaller ones. The first is that of *exponentiation*: given a cardinal κ, form the cardinal power 2^κ, which we know is always $>\kappa$. The second is that of *summation*: given a *set* of cardinals $\{\kappa_i : i \in I\}$, form the *sum* $\Sigma_{i \in I}\, \kappa_i$, which we know is always $>$ each κ_i. A notion of *largeness* which now suggests itself is the following. A cardinal is called *large* (in the present sense) if it is not obtainable from smaller cardinals by means of the operations of exponentiation and summation. More precisely, we define an infinite cardinal κ to be (strongly) *inaccessible* if (1) for every cardinal $\alpha < \kappa$ we have $2^\alpha < \kappa$ and (2) for each collection of cardinals $\{\alpha_i : i \in I\}$ with $\alpha_i < \kappa$ for every $i \in I$ and $|I| < \kappa$ (here $|I|$ denotes the cardinality of the set I), we have $\Sigma_{i \in I}\, \alpha_i < \kappa$. It is easy to see that the first infinite cardinal $\aleph_0$ is inaccessible; since the existence of this cardinal is ensured by the usual axioms of set theory it is customary to add a third proviso to the definition of an inaccessible cardinal κ, namely: (3) κ is *uncountable*.

It is clear that an inaccessible cardinal, if it exists, must be large in comparison with $\aleph_0$. For if κ is such a cardinal we must obviously have $\aleph_0 < \kappa$, $2^{\aleph_0} < \kappa$, $2^{2^{\aleph_0}} < \kappa$, $\aleph_0 + 2^{\aleph_0} + 2^{2^{\aleph_0}} + \ldots < \kappa$, *etc.* Moreover, it can be shown that if T is the collection of all sets X with the property that X, the members of X, the members of the members of X, *etc.* are all of cardinality $< \kappa$, then T is a model of the axioms of Zermelo–Fraenkel set theory (*ZF*), so by Gödel's second incompleteness theorem the existence of an inaccessible cardinal cannot be proved in *ZF* if this latter is consistent. It follows that the existence of an inaccessible cardinal is, so to speak, an article of faith whose truth is not guaranteed by the axioms of set theory. On the other hand, most set-theorists feel that the existence of such a cardinal (and even of a plurality of them) is a hypothesis hardly less plausible than the existence of $\aleph_0$. The 'justification' of this view frequently takes the following form. Starting with $\aleph_0$, consider the collection C of all cardinals obtainable by exponentiation and summation; the sum of C is then an inaccessible cardinal. (One can 'justify' the existence of $\aleph_0$ in the same way by starting with the cardinal 1!) Of course, this 'justifying' argument for the existence of inaccessibles cannot be carried out in *ZF* because the collection C cannot be proved to be a *set* in *ZF*. However, it *seems* reasonable to regard C as in some sense a well-defined totality, so the sum of C would also seem to be a perfectly acceptable cardinal. A similar argument, starting this time with arbitrary cardinal α, 'justifies' the existence of an inaccessible cardinal $\kappa > \alpha$. Accordingly, we are led to formulate the following two principles: (I) There exists an inaccessible cardinal; and its stronger version (II) for each cardinal α there is an inaccessible cardinal $\kappa > \alpha$. These two principles are examples of 'strong axioms of infinity'.

We have seen that principles (I) and (II) have natural justifications, but we also know that I (and hence II) is not provable in *ZF*. Thus the possibility arises that the existence of inaccessible cardinals might be *refutable* in *ZF*. Mathematical experience, however, appears to deny this possibility, and indeed it is hard to see how the 'justification' above could be shown to be inconsistent with *ZF*! On the other hand, Gödel's incompleteness theorem shows that one cannot hope for a convincing consistency proof for the existence of inaccessible cardinals relative to *ZF*. Thus the existence of inaccessible cardinals is truly an 'article of faith', albeit heuristically justifiable.

The strong axioms of infinity that we have just introduced are only the beginning. Much stronger (and less justifiable) principles are known: for example, that of the existence of *measurable cardinals*. To formulate this notion we need a definition. Let X be a set. A countably additive 2-valued *measure* on X is a function μ defined on the family of all subsets of X, taking values in the set $\{0, 1\}$ such that (*a*) $\mu(X) = 1$, (*b*) $\mu(\{x\}) = 0$ for all $x \in X$, and (*c*) if $\{X_n : n \in \omega\}$ is a countable disjoint family of subsets of X, then $\mu(\bigcup X_n) = \Sigma\mu(X_n)$. Banach and Ulam (in the 1930s) defined a cardinal κ to be *measurable* if there is a countably additive 2-valued measure on κ. (The notion derives from the classical idea of measure on the real line.) In 1939 Tarski showed that the first measurable cardinal (if it exists) is inaccessible. Very little else was known about measurable cardinals until 1960 when Hanf and Tarski showed that the first measurable cardinal κ must be an enormous inaccessible cardinal, in fact there must be κ inaccessibles *before* κ. More remarkable still was Scott's discovery in 1961 that the existence of a measurable cardinal contradicts Gödel's axiom of constructibility, which is to be contrasted with Gödel's earlier result that this axiom is consistent with the existence of an inaccessible cardinal, provided this latter is consistent with *ZF*. Scott's discovery shows that the presence of measurable cardinals affects not only the 'length' of the universe of sets, *i.e.* the 'number' of ordinals present, but also, so to speak its 'width'. Thus, the assumption of the existence of a measurable cardinal is a much more powerful (and risky) axiom of infinity than the mere postulation of the existence of an inaccessible cardinal. The term 'risky' is appropriate here because, unlike the situation for inaccessibles, there is no convincing justification for the existence of a measurable cardinal! In fact, some set-theorists suspect that the existence of such a cardinal might be *inconsistent* with *ZF*.

Despite these reservations, many set-theorists regard the existence of a measurable cardinal as a powerful new 'principle' of set theory, and have explored its consequences in great detail. For even if one regards the principle itself as suspect, its consequences may turn out to be independently verifiable (or at least plausible). And the existence of a measurable cardinal does have remarkable consequences! We have already mentioned Scott's result that, if there is a measurable cardinal, then there is a non-constructible set. Another is Rowbottom and Gaifman's result, that, under the same assumption, there is a non-constructible *real number*; in fact there are only *countably* many *constructible* real numbers. Still another is Martin's result that, if there is a measurable cardinal, then every Borel subset of the real line is determinate. This last is an example of a consequence of the assumption of the existence of measurable cardinals which was subsequently shown (by Martin) to be provable *without* the assumption, *i.e.* in $ZF+AC$. In this connection it is important to recall that in 1947 Gödel expressed the hope that a 'powerful new axiom' would be discovered which would decide the continuum hypothesis, *i.e.* the assertion that $2^{\aleph_0} = \aleph_1$. (Gödel rejected his own axiom of constructibility, which decides the continuum hypothesis in the affirmative, on the grounds that it is too restrictive.) In view of its remarkable effect on the continuum, one might hope that the assumption of the existence of measurable cardinals would meet Gödel's requirement. Unfortunately, this is not the case, for in 1967 Lévy and Solovay showed that, if the existence of a measurable cardinal is consistent with *ZF*, it is also consistent with the continuum hypothesis and with its negation (and later Silver showed it to be consistent with the generalised continuum hypothesis). In fact, although much more powerful axioms of infinity have been formulated since, not one of them decides the continuum hypothesis!

Drake's book contains a lucid account of the topics discussed above, and a lot more besides. The first three chapters provide a detailed development of Zermelo–Fraenkel set theory, leading to the Lévy hierarchy of formulas and the reflection principles. In chapter 4 we find a discussion of the classical results on inaccessible cardinals and their natural generalisation, Mahlo cardinals. Chapter 5 deals with the constructible universe and ordinal definable sets. In chapter 6 the author introduces measurable cardinals and the ultrapower construction, and proves the results of Scott and Rowbottom mentioned above. Chapters 7 and 8 are devoted to the theory of partition cardinals and to a proof of Silver and Solovay's important result that, if a partition cardinal exists, then there is a Δ^1_3 non-constructible real number. Chapter 9 deals with Scott and Hanf's notion of an indescribable cardinal, and chapter 10 with the relationship between large cardinals and languages with infinitely long formulas. The book contains many useful exercises and an excellent bibliography.

Drake's book seems to me outstanding in every respect. It is clearly written, well motivated, and contains a wealth of interesting detail which is never permitted to obscure the main thrust of the development. Furthermore, the author has had the courage to take a forthright philosophical stand with regard to the vexed question of the ontological status of the entities he discusses. He states in the preface that he has written the book 'from an uncompromisingly *realist* or *platonist* position; that is . . . that in some sense sets do exist, as objects to be studied, and that set theory is just as much about real objects as number theory'. (This position appears to be very close to Gödel's own.) However, with considerable fairness he goes on to say that 'I am well aware that other positions are taken by other mathematicians and I hope this book will be no less useful to them because of my foundational prejudices, which show through in many places; at least I have tried to make them open and explicit'. I for one am convinced that this book will be of great use to set-theorists, no matter what their philosophical position. One of my reasons for believing this is that the author's platonist viewpoint is in some sense very close to the original intuitive picture behind set theory and tends to direct attention to the 'sets themselves', rather than to a dense jungle of formalism. This is a very real advantage in a subject as technically complicated as the theory of large cardinals!

In conclusion, I can thoroughly recommend Drake's book *both* as an excellent reference volume *and* as an admirable text for any student of set theory wishing to become acquainted with some of the most important recent results in the subject.

JOHN BELL
London School of Economics

WOLLHEIM, R. (ed.) [1974]: *Freud, A Collection of Critical Essays.* New York: Doubleday. £3.00. Pp. viii+416.

Two abiding passions inform Freud's thought and writings: to understand the human heart and to set this understanding on a 'scientific' footing. Given his philosophical sensibilities, Freud found himself forced to take philosophical positions at every turn in the development of his views. The views themselves could not have been developed without facing philosophical problems. His contributions are, therefore, of intrinsic interest to the philosopher. It is important to be clear about Freud's philosophical presuppositions, to discuss them critically, and also to work through oneself the philosophical problems that faced him, if one is to appreciate his contribution. This kind of criticism is best represented in the present collection by Wittgenstein's 'Conversations on Freud' and Jerome Neu's long piece on 'Genetic Explanation in *Totem and Taboo*'. This task calls for intelligence, imagination and philosophical penetration. But there is another, complementary task, which is also essential if one is to cut through to Freud's insight into the human heart,[1] and this takes a special kind of sensitivity. Such sensitivity is best represented in the selection from Fingarette's book on *Self-deception* and Brian O'Shaughnessy's essay on 'The Id and the Thinking Process'.

---

[1] See Wisdom [1953], p. 261.

In its coverage this collection is fairly comprehensive. The essays vary in their philosophical approach and in the character of the sympathy they show for Freud's views. David Pears' essay is called 'Freud, Sartre and Self-deception', but its connection with Sartre is very tenuous and it contains nothing about Freud whatever. In fact the level of abstraction at which Pears operates and his intellectualism are hardly conducive to contact with what is at the heart of Freud's contribution. Peter Alexander, who at one level is 'in general sympathy with Freud's position', fares little better in this respect. I think that the excellent discussion of what Alexander says by Theodore Mischel shows this very clearly. Mischel's paper is short (pp. 322–31), but he does more than anyone I have read to bring out what Freud had in mind when he spoke of slips of the tongue in the way he did. His discussion of what it means for a patient to accept a psychoanalytic explanation of something he did, although equally brief, is also very good. Judging purely from what he says here he seems to be a philosopher who has learnt much from Wittgenstein. It is, however, interesting to note the difference in their relation to Freud when they write about him. What Wittgenstein says in the 'Conversations', though philosophically penetrating, has a certain ambivalence which would have prevented him from giving Freud a more sympathetic study, using his own important contributions to philosophical psychology to cut through to the more positive aspects of Freud's thought.

Stuart Hampshire, in his piece on 'Disposition and Memory', gives us an excellent example of this kind of work. He has obviously learnt much from Wittgenstein, yet what he says is the result of years of philosophical labour very much his own. In fact what we find in these pages (pp. 113–31) is a continuation of the work we find in his very interesting Inaugural Lecture at University College London in 1960 under the title 'Feeling and Expression'. For me 'Disposition and Memory' is the most interesting piece in the present collection.

Hampshire's essay divides into two parts. In the first part (pp. 113–20) he is concerned to elucidate the sense in which a person's mental states can exist independently of and in separation from the various expressions which they normally find in human life, thus constituting the texture of his inner life. This inner life, Hampshire argues, is a dimension which a child's life gradually acquires as he learns to identify his emotions and inclinations, to put them into words, and thereby to control them. This is made possible because these emotions and inclinations have natural expressions (Wittgenstein) which are in the first place constitutive of the emotions and inclinations themselves (p. 118). It enables their range to be extended (p. 115). More than this, it also extends the range within which he can exercise his will: 'Because the child may know what it is he is inclined to do, the question of whether he will do it comes up as a question for him to decide. He may *decide* to restrain his inclination to do something' (p. 115). 'His full inner life begins with his power of intentional inhibition' (p. 116). Hampshire's argument is that the use of language in the communication of emotions and inclinations and the inhibition of the actions they prompt are learnt in harness. The development of an inner life and the development of intentional states (the extention of the limits within which the will is operative so that it makes sense to speak of the 'freedom of the individual'[1]) are thus the two sides of the same coin.

In the second and more controversial part of his paper Hampshire begins by observing that the earliest behaviour of children already give hints of an inner life. He adds that the child has not as yet learned to avow what he feels and so he cannot be said to be aware of it. What he says here (top of p. 121) is not clear to me. He speaks of 'dispositions and inclinations that are repressed' and says that their 'signs', which cannot be intentional at this stage, have to be read by a third person. When he speaks of 'inhibited dispositions' here I take him to mean something akin to 'the inhibition of a trained animal'—*e.g.* 'of a dog trained to restrain its natural inclination to bite something' (p. 115). But in that case, how can Hampshire speak of an 'inner life' in the sense elucidated in the first part of his paper?

The gist of his answer is that when the adult remembers his childhood, the continuity that he discerns between past and present patterns, allows him to read present intentional expressions of feeling into the past patterns (p. 121). I do not find this answer satisfactory. Let me make this clear: I find it makes good sense for a person to look back on his past and see it in a different light from the one he saw it in at the time. In some such cases the continuity between the past and the present may justify us in saying that he was deceived, that his own mental state was other than he thought at the time, that his present apprehension of it is correct.[2] But in that case it must be possible for this person to have deceived himself, or at least to have been deceived, *at the time*. Yet how could this be if the time in question is a pre-verbal stage of his life? What I question, therefore, is whether at that stage an inclination is separable from its behavioural expression so as to constitute an element of the child's inner life. But if it isn't, can we speak of the 'unconscious' at that stage?

What Hampshire says about the notion of unconscious memories and the way such memories are operative in the present, influencing inclinations and conduct, is I think very good. I also think that Hampshire is absolutely right in finding a wide area of the influence of the past on the present which is not causal in character, but is a matter of the past continuing to live in the present and setting its seal on it. This is something that has been persistently misunderstood by most philosophers writing on Freud. Hence, for instance, going back to 1922, Professor Field asks: 'What I want to know is why the effect which the event produced on the person should be described as if the original event somehow went on existing and working?' The answer is that what is in question is not an event, but a thought or phantasy, nor merely one that the subject can recollect, but one which haunts him, though he does not recognise it. Surely this is the kind of thing that Freud has in mind when he writes:

> . . . the exciting experience continues in some way to be effective even years after, not indirectly by means of a chain of causes linking up with one another, but directly as the actual exciting cause—just as, for instance, the recollection of mental distress in full consciousness at some later period may stimulate tears (*Collected Papers*, vol. i, p. 29).

---

[1] Significantly this is the title of one of Hampshire's books.

[2] I have argued this in two different papers: Dilman [1959] and [1972].

This is the passage which ends with the famous words: 'hysterical patients suffer principally from reminiscences' (1892). Philosophers who have seen in Freud's abandonment of his early theories of seduction for the notion of traumatic phantasies nothing but a way of evading the possibility of refutation, have I think failed to appreciate that what we have here is an important step in the development of the Freudian conception of the unconscious and its importance in human life. What that importance amounts to, how it can restrict the sphere in which a man is capable of taking decisions or changing his mind, and how increasing the domain of consciousness and self-knowledge can restore the power to act differently—this is well discussed in the remaining part of Hampshire's paper—though necessarily the discussion is brief and coarse-grained. In a more detailed discussion it would be important to bring out the differences between ways in which one's past may hold one captive and ways in which recollection, re-living, and a reappraisal of one's present and past can bring increased freedom. Here one should beware of a one-sided diet of examples.

I now turn to the other contributions. Wollheim is right in wanting to connect 'identification' with 'imagination', but I think that he makes heavy weather of it in his philosophical argument. Also I think that he uses the notion of something 'internal' uncritically. He says that he 'should like to start by taking seriously a suggestion made by Wittgenstein . . . to the effect that, if we want to understand the nature of an inner phenomenon, we do well to consider an external counterpart of that phenomenon' (p. 181). But here 'internal' and 'external' should be in inverted commas. Wittgenstein wanted to combat the idea that one cannot give an adequate account of a mental phenomenon without referring to some internal episode or component. His criticism was that here one is looking in the wrong dimension, that one should turn one's attention to what he called 'the surrounding circumstances'. This is not to deny the existence of what I can keep to myself, but to claim that its character is logically dependent on its relation to what finds expression in human behaviour in particular circumstances. Wollheim's argument is that you cannot distinguish identification from imitation without bringing in imagination, and that doing so is in part to refer to something internal—almost as if to say that imitation lacks the inner dimension of identification. I agree that identification 'cannot be elucidated in exclusively behavioural terms' (p. 175). But it does not follow from this that it 'must be regarded as containing essentially an internal component' (*ibid.*). Why should not one say that identification is imitation in special circumstances—circumstances which remain to be elucidated? This is not to exclude a reference to the subject's thoughts and imaginings. But to understand what it means to have a thought or to imagine something we need to turn our attention to the language in which these are given expression and the kinds of activity in the weave of which this language is used. It is here that we shall find the logic peculiar to a thought or imagining, not in any additional component.[1]

The collection contains three papers on the question of the verification of psycho-analytic claims—general claims and particular interpretations. The first by Wesley Salmon was written in 1959. I find it philosophically crude and doctrinaire, though not devoid of penetration. In the first part Salmon argues that the 'principle of determinism' which Freud embraced is not the same as the kind of principle Hume resorts to in an attempt to 'justify induction', but is *a posteriori*. There is something right about this, though I cannot here bring out what it is.[1] In the second part of his paper Salmon deals with the charge that many psycho-analytic claims about what is unconscious are so formulated as to be immune to falsification. I believe what he says to be essentially correct. His view, very briefly, is that no 'few restricted items of behaviour can constitute conclusive evidence for or against the hypothesis . . . [it] must be judged on the weight of total evidence' (p. 280). He illustrates well what he means.

---

[1] See my [1968], p. 97.

Clark Glymour defends the view that 'there is a rational strategy for testing important parts of psycho-analysis . . . that relies almost exclusively on clinical evidence' (p. 287). 'Clinical' here is contrasted with 'experimental', and Glymour compares clinical testing with the kind of strategy that Kepler followed in testing his laws of planetary motion. J. O. Wisdom writes on 'Testing an Interpretation within a Session'. He formulates and examines the criteria used by psycho-analysts in judging whether the interpretations they give to their patients are correct. It is in the response of their patient to their interpretation that they look for confirmation. This raises at least two difficulties which J. O. Wisdom discusses: (*i*) that the response is brought about by the interpretation itself and may be the work of suggestion; (*ii*) that whether the response is or is not taken to confirm the interpretation depends on what one makes of it, which is itself a matter of interpretation, so that it looks as if one interpretation is tested by another. The paper is interesting and its author knows what he is writing about, but I do not find the way he meets these difficulties adequate.

I mentioned O'Shaughnessy's paper on 'The Id and the Thinking Process' which I like. There is a more general paper on 'Freud's Anatomies of the Self' by Irving Thalberg. It is a learned paper in which Thalberg traces the development and changes in Freud's conception of the divisions of the self, classifies the different analogies that dominated his thinking, and illustrates these with apt quotations. But despite the impressive subdivisions of the paper I find Thalberg's philosophical criticism a bit thin. Also the paper does little to promote an understanding of what is of positive value in Freud's thinking here, in the way that O'Shaughnessy does.

Ronald de Sousa's paper 'Norms and the Normal' is the one paper in this collection I do not know what to make of philosophically. David Sachs in his paper 'On Freud's Doctrine of Emotions' attributes to Freud the view that a person's emotional responses, however incongruous or disproportionate they may appear to be, can be seen never to be so in reality if one takes into account his thoughts, beliefs, conceptions, *etc.*, and the significance which things have for him. He tries to clarify this view and to show that if one appreciates that Freud held such a view, even though he may never have formulated it, this would give one 'a somewhat novel outlook upon his thought' (p. 133), namely a view that 'Freud's vision was a vision of the normalcy and health that, however obstructed or baffled, lie behind neurosis and psychosis'. The paper does raise interesting questions.

Fredric Weiss writes about 'Meaning and Dream Interpretation'. What he argues is in some ways similar to what Malcolm says in the Appendix of his book on *Dreaming* entitled 'Dreams and Psychiatry'. He expresses it in a kind of style I associate with Existentialism and Phenomenology: 'A dream-report is a peculiar object, in the respect that it may have a meaning-of its own, but it has no public meaning. If there is a meaning-of a dream-report, it is a meaning-for the subject. He gave it it; it is idiosycratic; it is something that he means' (p. 67).

---

[1] See my [1973].

Margaret Boden's paper is about Freudian Mechanisms of Defence and Computer Simulations of them. The largest part of the paper (section III) is concerned to give a clear factual account of such computer programmes. It is preceded (section II) by a short account of Freud's notion of 'defence mechanism' and followed (section IV) by an equally brief discussion of the limitations of such programmes. Her modest claim (if I understand her rightly) is very roughly this: You have to be very clear about the kind of thing you are writing a programme for a computer to simulate. Thus in writing and studying such programmes we shall be able to get a clear view about something we already know without being clear about its character and implications—something akin to conceptual clarification carried via a formalised language. But, of course, the greater the logical distance between the programme and what it is supposed to illuminate the dimmer will be the light it sheds on it. I am not acquainted with the work of K. M. Colby whose programmes Margaret Boden discusses, but I cannot say that I found much illumination in what she presents in her paper, though it is an able presentation and a clearly written and balanced paper.

The three papers by Thomas Nagel, Robert Solomon and John Marshall deal with Freud's early work in physiological psychology, its intrinsic philosophical interest, and its importance for understanding Freud's later more purely psychological contributions. Both Nagel and Solomon maintain that Freud was and remained a physicalist throughout the evolution of his thoughts. Both are concerned to defend this position as philosophically viable. Although they quote extensively from Freud's writings, it seems to me that both are more interested in the philosophical position illustrated in these passages than in Freud. In fact, I think that these passages, though certainly of intrinsic philosophical interest, are a dead-end for understanding Freud. I have personally found some of the things that Freud says in the *Project* and other early writings interesting in almost the opposite way, namely for containing the kernel of the kind of vision he was to develop in his maturity. In other words, I see the later Freud foreshadowed in his early writings, rather than the later Freud clinging to early philosophical positions. Insofar as these early philosophical positions remained with him—and they did—they were more of an encumbrance that prevented Freud from giving a clear expression to his insights. I believe that philosophy, if used with imagination, can help one to cut through this surface to the real content of Freud's contribution. It seems to me that in 'Freud's Anthropomorphism' Nagel's abstract philosophical preoccupations prevent him from gaining a proper understanding of Freud—*e.g.* his conception of the unconscious mind. In contrast, John Marshall's paper 'Freud's Psychology of Language' is more modest and also philosophically more illuminating.

Jerome Neu's paper on 'Genetic Explanation in *Totem and Taboo*' is a searching, thoughtful piece, in which he is patient with what he criticises and raises many penetrating questions.

To sum up, this philosophical collection on Freud contains twenty-one pieces on various aspects of Freud's contribution. They are all competent, though written from diverse philosophical points of view, and I have personally found six or seven of them really good and I learned from reading them—these, in the order in which they were arranged, are the pieces by Wittgenstein, Fingarette, Hampshire, O'Shaughnessy, Clark Glymour, Theodore Mischel and Jerome Neu.

İLHAM DILMAN
University College of Swansea

REFERENCES

DILMAN, I. [1959]: 'The Unconscious', *Mind*, **68**, pp. 446–73.
DILMAN, I. [1968]: 'Imagination', *Analysis*, **28**, pp. 90–7.
DILMAN, I. [1972]: 'The Unconscious: A Theoretical Construct?', *The Monist*, **56**, pp. 313–42.
DILMAN, I. [1973]: 'Freud and Psychological Determinism', *The Human World*, **11**, pp. 18–38.
WISDOM, J. [1953]: *Philosophy and Psycho-Analysis*.

PHILOSOPHY OF SCIENCE ASSOCIATION

Fifth Biennial Meeting

Time: 29–31 October 1976
Place: The Ambassador Hotels, Chicago, Illinois

The Philosophy of Science Association will hold its Fifth Biennial Meeting at The Ambassador Hotels, Chicago, Illinois, on 29–31 October 1976. The programme will include symposia on selected topics as well as sessions devoted to the presentation of contributed papers.

Those wishing to submit contributed papers should send *two copies* of the entire paper, with a 100-word summary abstract, to the Chairman of the Programme Committee, Professor Frederick Suppe, Committee on the History and Philosophy of Science, 1131 Skinner Building, University of Maryland, College Park, Maryland 20742, U.S.A. Closing dates for submission of papers is 1 February 1976. Papers on any topic within philosophy of science, including philosophical treatments of methodological or foundational issues in particular sciences, are eligible for consideration; and they may be written from any philosophical standpoint. They should be limited to a reading time of about 20 minutes (3,500 words). Although arrangements are being made to publish at least portions of the 1976 Biennial Meetings, acceptance of contributed papers for placement on the programme neither guarantees nor automatically precludes publication in the proceedings.

BRITISH SOCIETY FOR THE PHILOSOPHY OF SCIENCE

Annual Conference

The Society's 1976 Conference will be held at the University of Exeter on 24–26 September. The following sessions have been arranged:

Popper's World Three: Mr B. Carr (Exeter)
Discussant: Mr D. Miller (Warwick)
Piaget's Philosophy of Mind: Dr B. Rotman (Bristol)
Discussant: Dr D. Tall (Warwick)

Biology and Culture: Mr A. N. Maxwell (London)
Discussant: Mr A. Matus (London)

Consciousness and Neurology: Dr R. Hunter (London)
Discussant: Professor L. Hudson (Edinburgh)

The conference is being organised on behalf of the Society by Mr Brian Carr of the Department of Philosophy, University of Exeter, The Queen's Drive, Exeter, Devon, EX4 4QH, from whom further details may be obtained. (Please mark envelopes 'BSPS Conference'.)

LINGUISTICS AND PHILOSOPHY

An International Journal

Managing Editor: Robert Wall (University of Texas, Austin, Tex., U.S.A.).
Executive Editorial Committee: Max Cresswell (Victoria University, Wellington, New Zealand); Östen Dahl (Göteborg University, Sweden); Hans Kamp (University of London, England); Lauri Karttunen (University of Texas, Austin, Tex., U.S.A.); Asa Kasher (Tel-Aviv University, Israel); P. Stanley Peters, Jr. (University of Texas, Austin, Tex., U.S.A.).

Linguistics and Philosophy is a journal for studies focused on natural language which are of interest to both disciplines mentioned in the title. Although the area thus described is so extensive that a complete listing of relevant topics is precluded, at least the following specifically fall within it:

> traditional areas in the philosophy of language such as meaning and truth, reference, description, entailment, speech acts
>
> traditional areas of linguistics such as syntax, semantics and pragmatics (when the studies are of sufficient explicitness and generality to be also of philosophical interest)
>
> systems of logic with strong connections to natural language: modal logic, tense logic, epistemic logic, intentional logic
>
> philosophical questions raised by linguistics as a science: linguistic methodology, the status of linguistic theories, the nature of linguistic universals
>
> philosophically interesting problems at the intersection of linguistics and other disciplines: language acquisition, language and perception, language as a social convention

Contributions may be in the form of articles, review articles, notes, discussions, or remarks and replies. Reviews of books and monographs will be relatively few in number and restricted to those publications of the widest possible interest.

Articles for publication and communications relative to editorial matters should be sent to: Robert Wall, Department of Linguistics, University of Texas, Austin, Texas 78712, U.S.A. or to any member of the Executive Editorial Committee.

Two volumes will be published yearly. Subscription price per volume of 3 issues (approx. 450 pages) Dfl. 90.00/U.S. $36.00, plus Dfl. 12.50/U.S. $5.00 for postage and handling.

Private persons (no institutions, libraries, etc.) may subscribe at the reduced rate of Dfl. 42.50/U.S. $17.00 per volume, including postage and handling.

Subscriptions should be sent to: D. Reidel Publishing Company, P.O. Box 17, Dordrecht–Holland or D. Reidel Publishing Company, Inc., Lincoln Building, 160 Old Derby street, Hingham, Mass. 02043, U.S.A. or to any subscription agent.

Brit. J. Phil. Sci. 27 (1976), 201–223 *Printed in Great Britain*

A Renaissance of Empiricism in the Recent Philosophy of Mathematics[*]

by IMRE LAKATOS

Introduction
1 *Empiricism and Induction: the New Vogue in Mathematical Philosophy?*
2 *Quasi-Empirical versus Euclidean Theories.*
3 *Mathematics is Quasi-Empirical.*
4 *'Potential Falsifiers' in Mathematics.*
5 *Periods of Stagnation in the Growth of Quasi-Empirical Theories.*

INTRODUCTION

According to logical empiricist orthodoxy, while science is *a posteriori*, contentful and (at least in principle) fallible, mathematics is *a priori*, tautologous and infallible.[1] It may therefore come as a surprise for the historian of ideas to find statements by some of the best contemporary experts in foundational studies that seem to herald a renaissance of Mill's radical assimilation of mathematics to science. In the next section I present a rather long list of such statements. I then go on (in section 2) to explain the motivation and rationale of these statements. I then argue (in section 3) for what I call the 'quasi-empirical' nature of mathematics, as a whole. This presents a problem—namely what kind of statements may play the role of potential falsifiers in mathematics. I investigate this problem in section 4. Finally, in section 5, I examine briefly periods of stagnation in the growth of 'quasi-empirical' theories.

---

[*] This paper developed out of some remarks made in the course of a discussion at a conference in 1965. This discussion was initiated by a paper of Professor Kálmar's on an empirical approach to mathematics. (The whole discussion is published in Lakatos (*ed.*) [1967]).

The paper was originally accepted for publication in this *Journal* in 1967. Certain changes were suggested, however, and whilst making these Lakatos decided the paper needed further modifications. Only some of these had been made at the time of his death in February 1974.

Lakatos's general epistemological position changed considerably in the period 1967–74 (see particularly his [1970]). No doubt this would have induced substantial changes in this paper, especially in Sections 2, 4 and 5.

In preparing the paper for the press we have introduced a few minor changes in style and presentation. We have also added a few footnotes (indicated by a '*').

JOHN WORRALL and GREGORY CURRIE

[1] This empiricist position (and one of its central difficulties) is very clearly described by Ayer in his [1936]:

> . . . Whereas a scientific generalisation is readily admitted to be fallible, the truths of mathematics and logic appear to everyone to be necessary and certain. But if empiricism is correct no proposition which has a factual content can be necessary or certain. Accordingly the empiricist must deal with the truths of logic and mathematics in one of the two following ways: he must say either that they are not necessary truths, in which case he must account for the universal conviction that they are; or he must say that they have no factual content, and then he must explain how a proposition which is empty of all factual content can be true and useful and surprising. (pp. 72–3)

1 EMPIRICISM AND INDUCTION: THE NEW VOGUE IN MATHEMATICAL PHILOSOPHY?

Russell was probably the first modern logician to claim that the evidence for mathematics and logic may be 'inductive'. He, who in 1901 had claimed that the 'edifice of mathematical truths stands unshakable and inexpugnable to all the weapons of doubting cynicism',[1] in 1924 thought that logic (and mathematics) is exactly like Maxwell's equations of electrodynamics: both 'are believed because of the observed truth of certain of their logical consequences.'[2]

Fraenkel claimed in 1927 that:

> the intuitive or logical self-evidence of the principles chosen as axioms [of set theory] naturally plays a certain but not decisive role; some axioms receive their full weight rather from the self-evidence of the consequences which could not be derived without them.[3]

And he compared the situation of set theory in 1927 with the situation of the infinitesimal calculus in the eighteenth century, recalling d'Alembert's '*Allez en avant, et la foi vous viendra*'.[4]

Carnap, who at the 1930 conference in Königsberg still thought that 'any uncertainty in the foundations of the "most certain of all the sciences" is extremely disconcerting',[5] had decided by 1958 that there is an analogy—if only a distant one—between physics and mathematics: 'the impossibilty of absolute certainty'.[6]

Curry drew similar conclusions in 1963:

> The search for absolute certainty was evidently a principal motivation for both Brouwer and Hilbert. But does mathematics need absolute certainty for its justification? In particular, why do we need to be sure that a theory is consistent,

---

[1] Russell [1901], p. 57

[2] Russell [1924], pp. 325–6. He obviously hesitated between the view that one can put up with this state of affairs (and work out some sort of inductive logic for the *Principia*), and the view that one has to go on with the search for self-evident axioms. In the Introduction to the second edition of the *Principia*, he says that one *cannot* rest content with an axiom that has mere inductive evidence (p. xiv), while on p. 59 he devotes a little chapter to the 'Reasons for Accepting the Axiom of Reducibility' (although still not giving up the hope of deducing it from some self-evident truth).

[3] Fraenkel [1927], p. 61.

[4] Fraenkel, *op. cit.*, p. 61.

[5] Carnap [1931], p. 31. English translation in Benacerraf and Putnam (*eds.*) [1964].

[6] Carnap [1958], p. 240.

or that it can be derived by an absolutely certain intuition of pure time, before we use it? In no other science do we make such demands. In physics all theorems are hypothetical; we adopt a theory so long as it makes useful predictions and modify or discard it as soon as it does not. This is what has happened to mathematical theories in the past, where the discovery of contradictions had led to modifications in the mathematical doctrines accepted up to the time of that discovery. Why should we not do the same in the future? Using formalistic conceptions to explain what a theory is, we accept a theory as long as it is useful, satisfies such conditions of naturalness and simplicity as are reasonable at that time, and is not known to lead us into error. We must keep our theories under surveillance to see that these conditions are fulfilled and to get all the presumptive evidence of adequacy that we can. The Gödel theorem suggests that this is all we can do; an empirical philosophy of science suggests it is all we should do.[1]

To quote Quine:

> We may more reasonably view set theory, and mathematics generally, in much the way in which we view theoretical portions of the natural sciences themselves; as comprising truths or hypotheses which are to be vindicated less by the pure light of reason than by the indirect systematic contribution which they make to the organizing of empirical data in the natural sciences.[2]

And later he said:

> To say that mathematics in general has been reduced to logic hints at some new firming up of mathematics at its foundations. This is misleading. Set theory is less settled and more conjectural than the classical mathematical superstructure than can be founded upon it.[3]

Rosser too belongs to the new fallibilist camp:

> According to a theorem of Gödel . . . if a system of logic is adequate for even a reasonable facsimile of present-day mathematics, then there can be no adequate assurance that it is free from contradiction. Failure to derive the known paradoxes is very negative assurance at best and may merely indicate lack of skill on our part . . .[4]

Church, in 1939 thought that: 'there is no convincing basis for a belief in the consistency either of Russell's or of Zermelo's system, even as probable'.[5]

Gödel in 1944 stressed that under the influence of modern criticism of its foundations, mathematics has already lost a good deal of its 'absolute certainty' and that in the future, by the appearance of further axioms of set theory, it will be increasingly fallible.[6]

In 1947, developing this idea, he explained that for some such new axiom,

> even in case it had no intrinsic necessity at all, a (probable) decision about its truth is possible also in another way, namely, inductively by studying its 'success',

---

[1] Curry [1963], p. 16. See also his [1951], p. 61.
[2] Quine [1958], p. 4.
[3] Quine [1965], p. 125.
[4] Rosser [1953], p. 207.
[5] Church [1939].
[6] Gödel [1944], p. 213.

> that is, its fruitfulness in consequences demonstrable without the new axiom, whose proofs by means of the new axiom, however, are considerably simpler and easier to discover, and make it possible to condense into one proof many different proofs. The axioms for the system of real numbers, rejected by the intuitionists, have in this sense been verified to some extent owing to the fact that analytical number theory frequently allows us to prove number theoretical theorems which can subsequently be verified by elementary methods. A much higher degree of verification than that, however, is conceivable. There might exist axioms so abundant in their verifiable consequences, shedding so much light upon a whole discipline, and furnishing such powerful methods for solving given problems (and even solving them, as far as that is possible, in a constructivistic way) that quite irrespective of their intrinsic necessity they would have to be assumed at least in the same sense as any well established physical theory.[1]

Also, he is reported to have said a few years later that:

> the role of the alleged 'foundations' is rather comparable to the function discharged, in physical theory, by explanatory hypotheses . . . The so-called logical or set-theoretical 'foundation' for number-theory or of any other well established mathematical theory, is explanatory, rather than really foundational, exactly as in physics where the actual function of axioms is to *explain* the phenomena described by the theorems of this system rather than to provide a genuine 'foundation' for such theorems.[2]

Weyl says that non-intuitionistic mathematics can be tested, but not proved:

> No Hilbert will be able to assure us of consistency forever; we must be content if a simple axiomatic system of mathematics has met the test of our elaborate mathematical experiments so far . . . A truly realistic mathematics should be conceived, in line with physics, as a branch of the theoretical construction of the one real world, and should adopt the same sober and cautious attitude toward hypothetic extensions of its foundations as is exhibited by physics.[3]

Von Neumann, in 1947, concluded that

> After all, classical mathematics, even though one could never again be absolutely certain of its reliabilty, . . . stood on at least as sound a foundation as, for example, the existence of the electron. Hence, if one was willing to accept the sciences, one might as well accept the classical system of mathematics.[4]

Bernays argues very similarly: It is of course surprising and puzzling that the more content and power mathematical methods have, the less is their self-evidence. But 'this will not be so surprising if we consider that there are similar conditions in theoretical physics'.[5]

According to Mostowski mathematics is just one of the natural sciences:

---

[1] Gödel [1947], p. 521. The word 'probable' was inserted in the reprinted version, Gödel [1964], p. 265.

[2] Mehlberg [1962], p. 86.

[3] Weyl [1949], p. 235.

[4] Neumann [1947], pp. 189–90

[5] Bernays [1939], p. 83.

> [Gödel's] and other negative results confirm the assertion of materialistic philosophy that mathematics is in the last resort a natural science, that its notions and methods are rooted in experience and that attempts at establishing the foundations of mathematics without taking into account its originating in the natural sciences are bound to fail.[1]

And Kalmar agrees: '... the consistency of most of our formal systems is an empirical fact; ... Why do we not confess that mathematics, like other sciences, is ultimately based upon, and has to be tested in, practice?'[2]

These statements describe a genuine revolutionary turn in the philosophy of mathematics. Some describe their individual *volte-face* in dramatic terms. Russell, in his autobiography, says: 'The splendid certainty which I had always hoped to find in mathematics was lost in a bewildering maze ...'.[3] Von Neumann writes: 'I know myself how humiliatingly easily my own views regarding the absolute mathematical truth changed ..., and how they changed three times in succession!'[4] Weyl, recognising before Gödel that classical mathematics was *unrescuably* fallible, refers to this state of affairs as 'hard fact'.[5]

We could go on quoting; but surely this is enough to show that mathematical empiricism and inductivism (not only as regards the *origin* or *method*, but also as regards the *justification*, of mathematics) is more alive and widespread than many seem to think. But what is the background and what is the *rationale* of this new empiricist-inductivist mood? Can one give it a sharp, *criticisable* formulation?

2 QUASI-EMPIRICAL VERSUS EUCLIDEAN THEORIES

Classical epistemology has for two thousand years modelled its ideal of a theory, whether scientific or mathematical, on its conception of Euclidean geometry. The ideal theory is a deductive system with an indubitable truth-injection at the top (a finite conjuction of axioms)—so that truth, flowing down from the top through the safe truth-preserving channels of valid inferences, inundates the whole system.

It was a major shock for over-optimistic rationalism that science—in spite of immense efforts—could not be organised in such Euclidean theories. Scientific theories turned out to be organised in deductive systems where the *crucial* truth-value injection was *at the bottom*—at a special set of theorems. But *truth* does not flow upwards. The important logical flow in such *quasi-empirical theories* is not the transmission of truth but rather the retransmission of *falsity*—from special theorems at the bottom ('basic statements') up towards the set of axioms.[1]

---

[1] Mostowski [1955], p. 42.

[2] Kalmár [1967], pp. 192–3.

[3] Russell [1959], p. 212. For further details about Russell's turn, *cf.* my [1962].

[4] Neumann [1947], p. 190.

[5] Weyl [1928], p. 87.

Perhaps the best way to characterise quasi-empirical, as opposed to Euclidean theories, is this. Let us call those sentences of a deductive system in which some truth values are initially injected, 'basic statements', and the subset of basic statements which receive the particular value true, 'true basic statements'. Then a system is Euclidean if it is the *deductive closure* of those of its basic statements which are assumed to be true. Otherwise it is quasi-empirical.

An important feature of both Euclidean and quasi-empirical systems is the set of particular (usually unwritten) conventions regulating truth value injections in the basic statements.

A Euclidean theory may be claimed to be true; a quasi-empirical theory—at best—to be well-corroborated, but always conjectural. Also, in a Euclidean theory the true basic statements at the 'top' of the deductive system (usually called 'axioms') *prove*, as it were, the rest of the system; in a quasi-empirical theory the (true) basic statements are *explained* by the rest of the system.

Whether a deductive system is Euclidean or quasi-empirical is decided by the pattern of truth value flow in the system. The system is Euclidean if the characteristic flow is the transmission of truth from the set of axioms 'downwards' to the rest of the system—logic here is an *organon of proof;* it is quasi-empirical if the characteristic flow is retransmission of falsity from the false basic statements 'upwards' towards the 'hypothesis'—logic here is an *organon of criticism.*[2] But this demarcation between patterns of truth value flow is independent of the particular conventions that regulate the original truth value injection into the basic statements. For instance *a theory which is quasi-empirical in my sense may be either empirical or non-empirical in the usual sense:* it is empirical only if its basic theorems are spatio-temporally singular basic statements whose truth values are decided by the time-honoured but unwritten code of the experimental scientist.[3] (We may speak, even more generally, of Euclidean versus quasi-empirical theories independently of *what* flows in the logical channels: certain or fallible truth and falsehood, probability and improbability, moral desirability or undesirability, *etc. It is the how of the flow that is decisive.*)

The methodology of a science is heavily dependent on whether it aims at a Euclidean or at a quasi-empirical ideal. The basic rule in a science which adopts the former aim is to search for self-evident axioms—

---

[1] For an exposition of the story see my [1962]. The concept and term 'basic statement' is due to Karl R. Popper; see his [1934], ch. v.

[2] *Cf.* Popper [1963], p. 64.

[3] For a discussion *cf.* my [1971].

Euclidean methodology is puritanical, antispeculative. The basic rule of the latter is to search for bold, imaginative hypotheses with high explanatory and 'heuristic' power,[1] indeed, it advocates a proliferation of alternative hypotheses to be weeded out by severe criticism—quasi-empirical methodology is uninhibitedly speculative.[2]

The development of Euclidean theory consists of three stages: first the naive prescientific stage of trial and error which constitutes the prehistory of the subject; this is followed by the foundational period which reorganises the discipline, trims the obscure borders, establishes the deductive structure of the safe kernel; all that is then left is the solution of problems inside the system, mainly constructing proofs or disproofs of interesting conjectures. (The discovery of a decision method for theoremhood may abolish this stage altogether and put an end to the development.)

The development of a quasi-empirical theory is very different. It starts with problems followed by daring solutions, then by severe tests, refutations. The vehicle of progress is bold speculations, criticism, controversy between rival theories, problem-shifts. Attention is always focused on the obscure borders. The slogans are growth and permanent revolution, not foundations and accumulation of eternal truths.

The main pattern of Euclidean criticism is suspicion: Do the proofs really prove? Are the methods used too strong and therefore fallible? The main pattern of quasi-empirical criticism is proliferation of theories and refutation.

3 MATHEMATICS IS QUASI-EMPIRICAL

By the turn of this century mathematics, 'the paradigm of certainty and truth', seemed to be the last real stronghold of orthodox Euclideans. But there were certainly some flaws in the Euclidean organisation even of mathematics, and these flaws caused considerable unrest. Thus the central problem of all foundational schools was: 'to establish once and for all the certitude of mathematical methods'.[3] However, foundational studies unexpectedly led to the conclusion that a Euclidean reorganisation of mathematics as a whole may be impossible; that at least the richest mathematical theories were, like scientific theories, quasi-empirical. Euclideanism suffered a defeat in its very stronghold.

The two major attempts at a perfect Euclidean reorganisation of classical mathematics—logicism and formalism[1]—are well known, but a brief account of them from this point of view may be helpful.

---

[1] For the latter concept *cf.* Lakatos [1970].

[2] The elaboration of empirical methodology—which of course is the paradigm of quasi-empirical methodology—is due to Karl Popper.

[3] Hilbert [1925], p. 35.

(*a*) *The Frege–Russell approach* aimed to deduce all mathematical truths —with the help of ingenious definitions—from indubitably true logical axioms. It turned out that some of the logical (or rather set theoretical) axioms were not only not indubitably true but not even consistent. It turned out that the sophisticated second (and further) generations of logical (or set-theoretical) axioms—devised to avoid the known paradoxes—even if true, were not indubitably true (and not even indubitably consistent), and that the crucial evidence for them was that classical mathematics might be *explained*—but certainly not *proved*—by them.

Most mathematicians working on comprehensive '*grandes logiques*' are well aware of this. We have already referred to Russell, Fraenkel, Quine and Rosser. Their 'empiricist' turn is in fact a quasi-empiricist one: they realised (independently even of Gödel's results) that the *Principia Mathematica* and the strong set-theories, like Quine's *New Foundations* and *Mathematical Logic*, are all quasi-empirical.

Workers in this field are conscious of the method they follow: daring conjectures, proliferation of hypotheses, severe tests, refutations. Church's account of an interesting theory based on a restricted form of the law of excluded middle (later shown to be inconsistent by Kleene and Rosser[2]) outlines the quasi-empirical method:

> Whether the system of logic which results from our postulates is adequate for the development of mathematics, and whether it is wholly free from contradiction, are questions which we cannot now answer except by conjecture. Our proposal is to seek at least an empirical answer to these questions by carrying out in some detail a derivation of the consequences of our postulates, and it is hoped either that the system will turn out to satisfy the conditions of adequacy and freedom from contradiction or that it can be made to do so by modifications or additions.[3]

Quine characterised the crucial part of his *Mathematical Logic* as a 'daring structure . . . added at the constructor's peril'.[4] Soon it was shown by Rosser to be inconsistent and Quine then himself described his earlier characterisation as one that had 'a prophetic ring'.[1]

---

[1] Intuitionism is omitted: it never aimed at a reorganisation but at a truncation of classical mathematics. (**Editors' Note:* Not all the theorems of intuitionist mathematics are theorems of classical mathematics. In this sense, Lakatos is wrong to describe intuitionism as simply a 'truncation' of classical mathematics. Nevertheless, an important point remains. While Russell's logicism and Hilbert's formalism each regarded its task as the justification of the whole of classical mathematics, Brouwer's intuitionism was willing to jettison large parts of classical mathematics which do not meet its standards of justification.)

[2] Kleene and Rosser [1935].

[3] Church [1932], p. 348.

[4] Quine [1941*a*], p. 122. Some critics of Quine may say that it is only he who has made a 'daring' structure out of the natural simplicity of mathematics. But surely the Cantorian paradise is a 'bold theoretical construction, and as such the very opposite of analytical self-evidence' (Weyl [1947], p. 64). Also *cf.* the Weyl quotation in section 2.

One can never refute Euclideanism: even if forced to postulate highly sophisticated axioms, one can always stick to one's hopes of deriving them from some deeper layer of self-evident foundations.[2] There have been considerable and partly successful efforts to simplify Russell's *Principia* and similar logicistic systems. But while the results were mathematically interesting and important they could not retrieve the lost philosophical position. The *grandes logiques* cannot be proved true—nor even consistent; they can only be proved false—or even inconsistent.

(*b*) While the Frege–Russell approach aimed to turn mathematics into a unified classical Euclidean theory the *Hilbert approach* offered a radically new modification of the Euclidean programme, exciting both from the mathematical and the philosophical points of view.

Hilbertians claimed that classical analysis contains an absolutely true Euclidean kernel. But along side this there are 'ideal elements' and 'ideal statements' which, though indispensable for the deductive-heuristic machinery, are not absolutely true (in fact they are neither true nor false). But if the whole theory, containing both the concrete-*inhaltlich* and the ideal statements can be proved consistent in a Euclidean metamathematics,[3] the entire classical analysis would be saved.* That is, analysis *is* a quasi-empirical theory[4] but the Euclidean consistency proof will see to it that it should have no falsifiers. The sophistication of Cantorian speculation is to be safeguarded not by deeper-seated Euclidean axioms *in*

---

[1] Quine [1941*b*], p. 163. By the way, the most interesting feature of Rosser's paper is the search for ways of testing the consistency of *ML*. Rosser shows that 'if one can prove *201 from the remaining axioms, then the remaining axioms are inconsistent' (Rosser [1941], p. 97).

[2] Also, one can choose to cut down a quasi-empirical theory to its Euclidean kernel (that is the essential aspect of the intuitionist programme).

[3] Originally the metatheory was not to be axiomatised but was to consist of simple, protofinitary thought-experiments. In Bologna (1928) von Neumann even criticised Tarski for axiomatising it. (The generalisation of the concept of 'Euclidean theory' to informal, unaxiomatised theories does not constitute any difficulty.)

* *Editors' Note:* Lakatos is, perhaps, wrong to think that Hilbert's philosophy, at least as here presented, can be subsumed easily under Euclideanism. Metamathematics is an informal unaxiomatised theory and such theories do not have the required deductive structure to be candidates for Euclidean status. Informal theories can obviously be axiomatised, but one of Hilbert's central claims was that there was no need for this in the case of metamathematics (*cf.* footnote 2 *above*). Each principle assumed in a metamathematical proof was to be so obviously true as not to be in need of justification (or, rather, to be immediately justified by the so-called 'global intuition').

[4] To quote Weyl again:

> . . . whatever the ultimate value of Hilbert's program, his bold enterprise can claim one merit: it has disclosed to us the highly complicated and ticklish logical structure of mathematics, its maze of back-connections, which result in circles of which it cannot be gathered at a first glance whether they might not lead to blatant contradictions. (*op. cit.*, p. 61)

the theory itself—Russell has already failed in this venture—but by an austere Euclidean *metatheory*.

Eventually, Hilbertians defined the set of statements whose truth values could be regarded as directly given (the set of finitistically true statements) so clearly that their programme could be refuted.[1] The refutation was provided by Gödel's theorem which implied the impossibility of a finitary consistency proof for formalised arithmetic. The reaction of formalists is well summed up by Curry:

> This circumstance has led to a difference of opinion among modern formalists, or rather, it strengthened a difference of opinion which already existed. Some think that the consistency of mathematics cannot be established on *a priori* grounds alone and that mathematics must be justified in some other way. Others maintain that there are forms of reasoning which are *a priori* and constructive in a wider sense and that in terms of these the Hilbert program can be carried out.[2]

That is, either metamathematics was to be recognised as a quasi-empirical theory or the concept of finitary or *a priori* had to be stretched. Hilbert chose the latter option. According to him the class of *a priori* methods was now to include, for example, transfinite induction up to ϵ_0, used in Gentzen's proof of the consistency of arithmetic.

But not everybody was happy about this extension. Kálmar, who applied Gentzen's proof to the Hilbert–Bernays system, never believed that his proof was Euclidean. According to Kleene: 'To what extent the Gentzen proof can be accepted as securing classical number theory . . . is . . . a matter for individual judgment, depending on how ready one is to accept induction up to ϵ_0 as a finitary method.'[3] Or, to quote Tarski:

> . . . there seems to be a tendency among mathematical logicians to overemphasize the importance of consistency problems, and the philosophical value of the results so far in this direction seems somewhat dubious. Gentzen's proof of the consistency of arithmetic is undoubtedly a very interesting metamathematical result which may prove very stimulating and fruitful. I cannot say, however, that the consistency of arithmetic is now much more evident to me (at any rate, perhaps to use the terminology of the differential calculus, more evident than by epsilon) than it was before the proof was given. To clarify a little my reactions: let G be a formalism just adequate for formalizing Gentzen's proof, and let A be the formalism of arithmetic. It is interesting that the consistency of A can be proved in G; it would perhaps be equally interesting if it should turn out that the consistency of G can be proved in A.[4]

However, even those who find transfinite induction up to ϵ_0 infallible would not be happy to go on stretching the concept of infallibility so as to accommodate consistency proofs of stronger theories. In this sense 'the real test of proof-theory will be the proof of the consistency of *analysis*',[1] and this has still to be seen.

---

[1] Herbrand [1930], p. 248. It took three decades to arrive at this definition.

[2] Curry [1963], p. 11.

[3] Kleene [1952], p. 479.

[4] Tarski [1954], p. 19.

Gödel's and Tarski's incompleteness results however reduce the chances of the final success of Hilbert's programme still further. For if *extant* arithmetic cannot be proved by the original Hilbertian standards, the gradual, consistent (and indeed, ω-consistent) augmentation of theories containing arithmetic by further axioms can only be reached by still more fallible methods. That is, the future development of arithmetic will increase its fallibility. Gödel himself has pointed this out in his paper on Russell's mathematical logic:

> [Russell] compares the axioms of logic and mathematics with the laws of nature and logical evidence with sense perception, so that the axioms need not necessarily be evident in themselves, but rather their justification lies (exactly as in physics) in the fact that they make it possible for these 'sense perceptions' to be deduced; which of course would not exclude that they also have a kind of intrinsic plausibility similar to that in physics. I think that (provided 'evidence' is understood in a sufficiently strict sense) this view has been largely justified by subsequent developments, and it is to be expected that it will be still more so in the future. It has turned out that (under the assumption that modern mathematics is consistent) the solution of certain arithmetical problems requires the use of assumptions essentially transcending arithmetic, *i.e.*, the domain of the kind of elementary indisputable evidence that may be most fittingly compared with sense perception. Furthermore it seems likely that for deciding certain questions of abstract set theory and even for certain related questions of the theory of real numbers new axioms based on some hitherto unknown idea will be necessary. Perhaps also the apparently unsurmountable difficulties which some other mathematical problems have been presenting for many years are due to the fact that the necessary axioms have not yet been found. Of course, under these circumstances mathematics may lose a good deal of its 'absolute certainty'; but, under the influence of the modern criticism of the foundations, this has already happened to a large extent. There is some resemblance between this conception of Russell and Hilbert's 'supplementing the data of mathematical intuition' by such axioms as, e.g., the law of excluded middle which are not given by intuition according to Hilbert's view; the borderline however between data and assumptions would seem to lie in different places according to whether we follow Hilbert or Russell.[2]

Quine says that in the field of *grande logique* construction 'at the latest, the truism idea received its deathblow from Gödel's incompleteness theorem. Gödel's incompleteness theorem can be made to show that we can never approach *completeness* of elementhood axioms without approaching contradiction . . .'[3]

---

[1] Bernays and Hilbert [1939], p. vii.

[2] Gödel [1944], p. 213.

[3] Quine [1941*a*], p. 127.

There are many possible ways of augmenting systems including arithmetic. One is through adding strong, arithmetically testable, axioms of infinity to *grande logiques*.[1] Another is through constructing strong ordinal logics.[2] A third one is to allow non-constructive rules of inference.[3] A fourth one is the model-theoretic approach.[4] But all of them are fallible, not less fallible—and not less quasi-empirical—than the ordinary classical mathematics which was so much in want of foundations. This recognition—that not only the *grandes logiques*, but also mathematics is quasi-empirical—is reflected in the 'empiricist' statements by Gödel, von Neumann, Kalmar, Weyl and others.

(It should however be pointed out that some people believe that some of the principles used in these different methods are *a priori* and they were arrived at by 'reflection'. For instance, Gödel's empiricism is qualified by the hope that set theoretical principles may be found which are *a priori* true. He claims that Mahlo's 'axioms show clearly, not only that the axiomatic system of set theory as used today is incomplete, but also that it can be supplemented without arbitrariness by new axioms which only unfold the content of the concept of set explained above'.[5] (Gödel, however, does not seem to be very sure of the *a priori* characterisability of the concept of set, as is evident from his already quoted quasi-empiricist remarks and also from his hesitation in his [1938], where he says that the axiom of constructibility 'seems to give a natural completion of the axioms of set theory, in so far as it determines the vague notion of an arbitrary infinite set in a definite way'.[6]) Weyl actually made fun of Gödel's over-optimistic stretching of the possibilities of *a priori* knowledge:

> Gödel, with his basic trust in transcendental logic, likes to think that our logical optics is only slightly out of focus and hopes that after some minor correction of it we shall see *sharp*, and then everybody will agree that we see *right*. But he who does not share this trust will be disturbed by the high degree of arbitrariness involved in a system like *Z*, or even in Hilbert's system. How much more convincing and closer to facts are the heuristic arguments and the subsequent systematic constructions in Einstein's general relativity theory, or the Heisenberg

---

[1] Such strong axioms were formulated by Mahlo, Tarski and Levy. As to the arithmetical testability of these axioms:

> It can be proved that these axioms also have consequences far outside the domain of very great transfinite numbers, which is their immediate subject matter: each of them, under the assumption of its consistency, can be shown to increase the number of decidable propositions even in the field of Diophantine equations. (Gödel [1947]. p. 520)

[2] This line of research was initiated by Turing ([1939]) and developed by Feferman ([1968]).

[3] *Cf. e.g.* Rosser [1937]; Tarski [1939]; Kleene [1943].

[4] *Cf.* Kemeny [1958], p. 164.

[5] Gödel [1964], p. 264 (*Cf.* Gödel [1947], p. 520).

[6] Gödel [1938], p. 557.

> –Schrödinger quantum mechanics. A truly realistic mathematics should be conceived, in line with physics, as a branch of the theoretical construction of the one real world, and should adopt the same sober and cautious attitude towards hypothetic extensions of its foundations as is exhibited by physics.[1]

Kreisel, however, extols this sort of aprioristic reflection by which, he claims, one gains set theoretical axioms, and 'right' definitions, and calls anti-apriorism an 'antiphilosophic attitude' and the idea of progress by trial and error empirically false.[2] What is more, in his reply to Bar–Hillel, he wants to extend this method to science, thereby re-discovering Aristotelian essentialism. He adds: 'If I were really convinced that reflection is extraordinary or illusory I should certainly not choose philosophy as a profession; or, having chosen it, I'd get out fast'.[3] In his comment on Mostowski's paper he tries to play down Gödel's hesitation as out of date.[4] But just as Gödel immediately refers to inductive evidence, Kreisel refers (in the Reply) to the 'limitations' of the heuristic of reflection. (So, after all, 'reflection', 'explication' *is* fallible.)

4 THE POTENTIAL FALSIFIERS OF MATHEMATICS

If mathematics and science are both quasi-empirical, the crucial difference between them, if any, must be in the nature of their 'basic statements', or 'potential falsifiers'. The 'nature' of a quasi-empirical theory is decided by the nature of the truth-value injections into its potential falsifiers.[5] Now nobody will claim that mathematics is empirical in the sense that its potential falsifiers are singular spatio-temporal statements. But then what is the nature of mathematics? Or, what is the nature of the potential falsifiers of mathematical theories?[6] The very question would have been an insult in the years of intellectual honeymoon of Russell or Hilbert. After all, the *Principia* or the *Grundlagen der Mathematik* were meant to put an end—once and for all—to counterexamples and refutations in mathematics. Even now the question still raises some eyebrows.

But comprehensive axiomatic set theories and systems of metamathematics, can be, and indeed have been, refuted. Let us first take comprehensive axiomatic set theories. Of course, they have *potential logical falsifiers:* statements of the form $p \mathbin{\&} \neg p$. But are there other falsifiers? The potential falsifiers of science, roughly speaking, express the 'hard facts'. But is there anything analogous to 'hard facts' in mathematics? If we accept the view that a formal axiomatic theory implicitly defines its subject-matter, then there would be no mathematical falsifiers except the logical ones.

---

[1] Weyl, *op. cit.*, p. 235.
[2] Kreisel [1967*a*], p. 140.
[3] Kreisel [1967*b*], p. 178.
[4] Kreisel [1967*c*], pp. 97–8.
[5] See *above*, p. 206.
[6] It is hoped that this Popperian formulation of the age-old question will shed new light on some questions in the philosophy of mathematics.

But if we insist that a formal theory should be the formalisation of some informal theory, then a formal theory may be said to be 'refuted' if one of its theorems is negated by the corresponding theorem of the informal theory. One could call such an informal theorem a ***heuristic falsifier*** of the formal theory.[1]

Not all formal mathematical theories are in equal danger of heuristic refutation in a given period. For instance, ***elementary group theory*** is scarcely in any danger: in this case the original informal theories have been so radically replaced by the axiomatic theory that heuristic refutations seem to be inconceivable.

Set theory is a subtler question. Some argue that after the total destruction of naive set theory by *logical* falsifiers one cannot speak any more of set-theoretical facts: one cannot speak of an ***intended*** interpretation of set theory any more. But even some of those who dismiss set-theoretical intuition may still agree that axiomatic set theories perform the task of being the dominant, unifying theory of mathematics in which all available mathematical facts (*i.e.* some specified subset of informal theorems) have to be explained. But then one can criticise a set theory in two ways: its axioms may be tested for consistency and its definitions may be tested for the 'correctness' of their translation of branches of mathematics like arithmetic. For instance, we may some day face a situation where some machine churns out a formal proof in a formal set theory of a formula whose intended meaning is that there exists a non-Goldbachian even number. At the same time a number theorist might prove informally that all even numbers are Goldbachian. If his proof can be formalised within our system of set-theory, then our theory will be inconsistent. But if it cannot be thus formalised, the formal set theory will not have been shown to be inconsistent, but only to be a *false* theory of arithmetic (while still being possibly a true theory of some mathematical structure that is not isomorphic with arithmetic). Then we may call the informally proved Goldbach theorem a ***heuristic falsifier***, or more specifically, an ***arithmetical falsifier*** of our formal set theory.[2] The formal theory is false in respect of the informal *explanandum* that it had set out to explain; we have to replace it by a better one. First we may try piecemeal improvements. It may have been only the definition of 'natural number' that went wrong and then the definition could be 'adjusted' to each heuristic falsifier. The axiomatic system itself (with its formation and transformation rules) would become useless as an explanation of arithmetic only if it was altogether 'numerically insegregative',[1] *i.e.* if it turned out that no finite sequence of adjustments of the definition eliminates *all* heuristic falsifiers.

---

[1] It would be interesting to investigate how far the demarcation between logical and heuristic falsifiers corresponds to Curry's demarcation between mathematical truth and 'quasi-truth' (or 'acceptability'). *Cf.* his [1951], especially ch. xi. Curry calls his philosophy 'formalist' as opposed to '*inhaltich*' or 'contensive' philosophies, like Platonism or intuitionism (Curry [1965], p. 80). However, besides his philosophy of formal structure, he has a philosophy of acceptability—but surely one cannot explain the growth of formal mathematics without acceptability considerations, so Curry offers an '*inhaltlich*' philosophy after all.

[2] The expression 'ω-consistency', is as Quine pointed out (Quine [1953] p. 117.), misleading. A demonstration of the 'ω-inconsistency' of a system of arithmetic would in fact be a ***heuristic*** falsification of it. Ironically, the historical origin of the misnomer was that the phenomenon was used by Gödel and Tarski precisely to divorce truth ('ω-consistency') from consistency.

Now the problem arises: *what class of informal theorems should be accepted as arithmetical falsifiers of a formal theory containing arithmetic?*

Hilbert would have accepted only finite numerical equations (without quantifiers) as falsifiers of formal arithmetic. But he could easily show that *all* true finite numerical equations are provable in his system. From this it followed that his system was complete with regard to true basic statements, therefore, if a theorem in it could be proved false by an arithmetical falsifier, the system was also inconsistent, for the formal version of the falsifier was already a theorem of the system. Hilbert's reduction of falsifiers to logical falsifiers (and thereby the reduction of truth to consistency) was achieved by a very narrow ('finitary') definition of arithmetical basic statements.

Gödel's informal proof of the truth of the Gödelian undecidable sentence posed the following problem: is the *Principia* or Hilbert's formalised arithmetic—on the assumption that each is consistent—true or false if we adjoin to it the negation of the Gödel sentence? According to Hilbert the question should have been meaningless, for Hilbert was an instrumentalist with regard to arithmetic outside the finitary kernel and would not have seen any difference between systems of arithmetic with the Gödel sentence or with its negation as long as they both equally implied the true basic statements (to which, by the way, his implicit meaning-and-truth-definition was restricted). Gödel proposed[2] to extend the range of (meaningful and true) basic statements from finitary numerical equations also to statements with quantifiers and the range of proofs to establish the truth of basic statements from 'finitary' proofs to a wider class of intuitionistic methods. It was this methodological proposal that divorced truth from consistency and introduced a new pattern of conjectures and refutations based on arithmetical falsifiability: it allowed for daring speculative theories with very strong, rich axioms while criticising them from the outside by informal theories with weak, parsimonious axioms. *Intuitionism is here used not for providing foundations but for*

---

[1] See Quine, *loc. cit.*, p. 118.

[2] See his intervention in 1930 in Königsberg; recorded in Gödel [1931].

providing falsifiers, not for discouraging but for encouraging and criticising speculation!

It is surprising how far constructive and even finite falsifiers can go in testing comprehensive set theories. Strong axioms of infinity for instance are testable in the field of Diophantine equations.[1]

But comprehensive axiomatic set theories do not have only arithmetical falsifiers. They may be refuted by theorems—or axioms—of naive set theory. For instance Specker 'refuted' Quine's *New Foundations* by proving in it that the ordinals are not well-ordered by '$\leqslant$' and that the axiom of choice must be given up.[2] Now *is* this 'refutation' of the *New Foundations*, even a heuristic refutation? Should the well-ordering theorem of shattered naive set theory overrule Quine's system? Even if, with Gödel and Kreisel, we consider naive set theory as re-established by Zermelo's correction,[3] we could admit the well-ordering theorem and the axiom of choice as a heuristic falsifier only if we again extend the class of (intuitionistic) heuristic falsifiers to (almost?) *any* theorem in corrected naive set theory. (We may call the former the class of *strong heuristic falsifiers* and the latter the class of *weak heuristic falsifiers*.) But this would surely be irrational: at best we have to consider them as two rival theories (*strictly* speaking *no* heuristic falsifier can be more than a rival hypothesis). After all nothing prevents us from forgetting about naive sets and focusing our attention on the new unintended model of *New Foundations*![4]

Indeed, we can go even further. For instance, if it turned out that all strong set-theoretical systems are arithmetically false, we may modify our arithmetic—the new, non-standard arithmetic may possibly serve the empirical sciences just as well. Rosser and Wang, who—three years before Specker's result—showed that in no model of *New Foundations* does '$\leqslant$' well-order both finite cardinals and infinite ordinals as long as we stick to the intended interpretation of '$\leqslant$', discuss this possibility:

> One may well question whether a formal logic which is known to have no standard model is a suitable framework for mathematical reasoning. The proof of the pudding is in the eating. For topics in the usual range of classical mathematical analysis, the reasoning procedures of Quine's *New Foundations* are as close to the accepted classical reasoning procedures as for any system known to us. However, in certain regions, notably when dealing with extremely large ordinals, the reasoning procedures of Quine's *New Foundations* reflect the absence of a standard model, and appear strange to the classically minded mathematician. However, since the theory of ordinals is suspect when applied

---

[1] See *above*, p. 212, fn. 1

[2] Specker [1953]; also *cf*. Quine [1963], p. 294 ff.

[3] *Cf*. Gödel [1947], p. 518 and Kreisel [1967].

[4] For philosophers of science after Popper it should anyway be a commonplace that *explanans* and *explanandum* may be rival hypotheses.

> to very large ordinals, it is hardly a serious defect in a logic if it makes this fact apparent.
>
> We suspect that the idea that a logic must have a standard model if it is to be acceptable as a framework for mathematical reasoning is merely a vestige of the old idea that there is such a thing as absolute mathematical truth. Certainly the requirements on a standard model are that it reflect certain classically conceived notions of the structure of equality, integers, ordinals, sets, *etc.* Perhaps these classically conceived notions are incompatible with the procedures of a strong mathematical system, in which case a formal logic for the strong mathematical system could not have a standard model.[1]

This of course amounts to the claim that the only real falsifiers are logical ones. But other mathematicians, Gödel for example, would surely reject the *New Foundations* on Specker's refutation: for him the axiom of choice and the well-ordering of ordinals are self-evident truths.[2]

No doubt the problem of basic statements in mathematics will attract increasing attention with the further development of comprehensive set theories. Recent work indicates that some very abstract axioms may soon be found testable in most unexpected branches of classical mathematics: *e.g.* Tarski's axiom of inaccessible ordinals in algebraic topology.[3] The continuum hypothesis also will provide a testing ground: the accumulation of further intuitive evidence against the continuum hypothesis may lead to the rejection of strong set theories which imply it.[3] Gödel [1964] enumerates quite a few implausible consequences of the continuum hypothesis: a crucial task of his new Euclidean programme is to provide a self-evident set theory from which its negation is derivable.[4]

If one regards comprehensive set theories—and mathematical theories in general—as quasi-empirical theories, a host of new and interesting problems arise. Until now the main demarcation has been between the proved and the unproved (and the provable and unprovable); radical justificationists ('Positivists') equated this demarcation with the demarcation between meaningful and meaningless. But now there will be a new demarcation problem: *the problem of demarcation between testable and untestable* (*metaphysical*) *mathematical theories with regard to a given set of basic statements.* Certainly one of the surprises of set theory was the fact that theories about sets of very high cardinality are testable in respect to a relatively modest kernel of basic statements (and thus have arithmetical content).[1] Such a criterion will be interesting and informative—but it would be unfortunate if some people should want to use it again as a meaning criterion as happened in the philosophy of science.

---

[1] Rosser and Wang [1950], p. 115.

[2] In his original paper [1947], Gödel says that the axiom of choice is exactly as evident as the other axioms 'in the present state of our knowledge' (p. 516). In the 1964 reprint (Gödel [1964]) this has been replaced by 'from almost every possible point of view' (p. 259 n. 2). He proposed, after some hesitation, a further extension of the range of set-theoretical basic statements that in fact amounted to a new Euclidean programme—but immediately proposed a quasi-empirical alternative in the case of failure. (See especially the supplement to his [1964].)

[3] *Cf.* Myhill [1960], p. 464.

[4] Kreisel criticises Gödel (Kreisel [1967*a*]) for not discussing his turn from proposing the constructibility axiom as a completion of set theory in 1938 to surreptitiously withdrawing it in 1947. One would think the reason for the turn is obvious: in the meantime he must have studied the work done on the consequences of the continuum hypothesis (mainly by Lusin and Sierpinski) and must have come to the conclusion that a set theory in which the hypothesis is deducible (like the one he suggested in 1938) is false. It may be interesting to note that according to Lusin a simple proposition in the theory of analytic sets which Sierpinski showed to be incompatible with the continuum hypothesis is 'indubitably true'—indeed he puts forward an impressive argument (Lusin [1935] and Sierpinski [1935]).

Another problem is that testability in mathematics rests on the slippery concept of heuristic falsifier. A heuristic falsifier after all is a falsifier only in a Pickwickian sense: it does not falsify the hypothesis, it only suggests a falsification—and suggestions can be ignored. It is only a rival hypothesis. But this does not separate mathematics as sharply from physics as one may think. Popperian basic statements too are only hypotheses after all. *The crucial role of heuristic refutations is to shift problems to more important ones*, to stimulate the development of theoretical frameworks with more content. One can show of most classical refutations in the history of science and mathematics that they are heuristic falsifications. The battle between rival mathematical theories is most frequently decided also by their relative explanatory power.[2]

Let us finally turn to the question: *what is the 'nature' of mathematics*, that is, on what basis are truth values injected into its potential falsifiers? This question can be in part reduced to the question: What is the nature of *informal* theories, that is, what is the nature of the potential falsifiers of *informal* theories? Are we going to arrive, tracing back problem-shifts through informal mathematical theories to empirical theories, so that mathematics will turn out in the end to be *indirectly empirical*, thus justifying Weyl's, von Neumann's and—in a certain sense—Mostowski's and Kalmar's position? Or is *construction* the only source of truth to be injected into a mathematical basic statement? Or *platonistic intuition*? Or *convention?* The answer will scarcely be a monolithic one. Careful historico-critical case-studies will probably lead to a sophisticated and composite solution. But whatever the solution may be, the naive school concepts of static rationality like *apriori-aposteriori*, *analytic-synthetic* will only hinder its emergence. These notions were devised by classical epistemology to classify Euclidean certain knowledge—for the problem-shifts in the growth of quasi-empirical knowledge they offer no guidance.*

---

[1] The term 'content' is here used in a Popperian sense: the 'arithmetical content' is the set of arithmetical falsifiers.

[2] *Cf.* Lakatos [1977]. **Editors' Note:* This study in rival research programmes in the theory of the continuum will be published in a volume of collected essays.

5 PERIODS OF STAGNATION IN THE GROWTH OF QUASI-EMPIRICAL THEORIES

This history of quasi-empirical theories is a history of daring speculations and dramatic refutations. But new theories and spectacular refutations (whether logical or heuristic) do not happen every day in the life of quasi-empirical theories, whether scientific or mathematical. There are occasional long *stagnating periods* when a single theory dominates the scene without having rivals or acknowledged refutations. Such periods make many forget about the criticisability of the basic assumptions. Theories, which looked counterintuitive or even perverted when first proposed, assume authority. Strange methodological delusions spread: some imagine that the axioms themselves start glittering in the light of Euclidean certainty, others imagine that the deductive channels of elementary logic have the power to retransmit truth (or probability) 'inductively' from the basic statements to the extant axioms.

The classical example of an abnormal period in the life of a quasi-empirical theory is the long domination of Newton's mechanics and theory of gravitation. The theory's paradoxical and implausible character put Newton himself into despair: but after a century of corroboration Kant thought it was self-evident. Whewell made the more sophisticated claim that it had been solidified by 'progressive intuition',[1] while Mill thought it was inductively proved.

Thus we may name these two delusions 'the Kant–Whewell delusion', and the 'inductivist delusion'. The first reverts to a form of Euclideanism;

---

* *Editors' note:* Since this paper was written a good deal of further work has been done on testing proposed set theoretical axioms, like the continuum hypothesis and strong axioms of infinity. (A good survey is to be found in Fraenkel, Bar Hillel and Levy [1973]. See also Shoenfield [1971] for the axiom of measurable cardinals.) Levy and Solovay's work ([1967]) indicates that large cardinal axioms will not decide the continuum problem. As another line of attack, alternatives to the continuum hypothesis have been formulated and tested. An example is 'Martin's axiom', which is a consequence of the continuum hypothesis, but consistent with its negation (see Martin and Solovay [1970] and Solovay and Tennenbaum [1971]). Of the six consequences of the Continuum Hypothesis which Gödel regarded as highly implausible, three follow also from Martin's Axiom. But Martin and Solovay take a different attitude to that taken by Gödel. They have, they say, 'virtually no intuitions' about the truth or falsity of these three consequences.

Lakatos does not, by the way, distinguish between two different types of mathematical consequences of axioms of this kind. Many of these consequences, for example about constructible real numbers, despite being about 'ordinary' mathematical objects are of interest only to set theoreticians. Few such consequences will be of a kind which are testable against informally proved mathematical theorems.

[1] *E.g.* Whewell [1860], especially ch. xxix.

the second establishes a new—inductivist—ideal of deductive theory where the channels of deduction can also carry truth (or some quasi-truth like probability) upwards, from the basic statements to the axioms.

The main danger of both delusions lies in their methodological effect: both trade the challenge and adventure of working in the atmosphere of permanent criticism of quasi-empirical theories for the torpor and sloth of a Euclidean or inductivist theory, where axioms are more or less established, where criticism and rival theories are discouraged.[1]

The gravest danger then in modern philosophy of mathematics is that those who recognise the fallibility and therefore science-likeness of mathematics, turn for analogies to a wrong image of science. The twin delusions of 'progressive intuition' and of induction can be discovered anew in the works of contemporary philosophers of mathematics.[2] These philosophers pay careful attention to the degrees of fallibility, to methods which are *a priori* to some degree, and even to degrees of rational belief. But scarcely anybody has studied the possibilities of refutations in mathematics.[3] In particular, nobody has studied the problem of how much of the Popperian conceptual framework of the logic of discovery in the empirical sciences is applicable to the logic of discovery in the quasi-empirical sciences in general and in mathematics in particular. *How can one take fallibilism seriously without taking the possibility of refutations seriously?* One should not pay lip-service to fallibilism: 'To a philosopher there can be nothing which is absolutely self-evident' and then go on to state: 'But in practice there are, of course, many things which can be called self-evident . . . each method of research presupposes certain results as self-evident.'[4] Such *soft fallibilism* divorces fallibilism from criticism and shows how deeply ingrained the Euclidean tradition is in mathematical philosophy. It will take more than the paradoxes and Gödel's results to prompt philosphers to take the empirical aspects of mathematics seriously, and to elaborate a philosophy of critical fallibilism, which takes inspiration not from the so-called foundations but from the *growth* of mathematical knowledge.

---

[1] *Cf.* Kuhn, especially his [1963].

[2] The main protagonists of Whewellian progressive intuition in mathematics are Bernays, Gödel, and Kreisel (see *above*, pp. 212–13). Gödel also provides an inductivist criterion of truth, should progressive (or as Carnap would call it 'guided') intuition fail: an axiomatic set theory is true if it is richly verified in informal mathematics or physics. 'The simplest case of an application of the criterion under discussion arises when some set-theoretical axiom has number-theoretical consequences verifiable by computation up to any given integer' (Supplement to Gödel [1964], p. 272).

[3] Kalmar—with his criticism of Church's thesis—is a notable exception (see Kalmar [1959]).

[4] Bernays [1965], p. 127

REFERENCES

(Where reprints are cited the page references are to these)

Ayer, A. J. [1936]: *Language, Truth and Logic*. London: Gollancz.

Bar Hillel, Y. [1967]: 'Is Mathematical Empiricism Still Alive?', in I. Lakatos (*ed.*): [1967], pp. 197–9.

Bernays, P. [1939]: 'Bemerkungen zur Grundlagenfrage', in F. Gonseth (*ed.*): *Philosophie Mathématique*, pp. 83–7. Paris: Hermann.

Bernays, P. [1965]: 'Some Empirical Aspects of Mathematics', in P. Bernays and S. Dockx (*eds.*): *Information and Prediction in Science*, pp. 123–8. New York: Academic Press.

Bernays, P. [1967]: 'Mathematics and Mental Experience', in I. Lakatos (*ed.*): [1967], pp. 196–7.

Bernays, P. and Hilbert, D. [1939]: *Grundlagen der Mathematik*, **2**. Berlin: Springer.

Carnap, R. [1931]: 'Die Logizistische Grundlegung der Mathematik', *Erkenntnis*, **2**, pp. 91–105. English translation in P. Benacerraf and H. Putnam (*eds.*): *Philosophy of Mathematics, Selected Readings*, pp. 31–41. Oxford: Basil Blackwell.

Carnap, R. [1958]: 'Beobachtungssprache und Theoretische Sprache', *Dialectica*, **12**, pp. 236–47.

Church, A. [1932]: 'A Set of Postulates for the Foundation of Logic', *Annals of Mathematics*, **33**, Second Series, pp. 346–66.

Church, A. [1939]: 'The Present Situation in the Foundations of Mathematics', in F. Gonseth (*ed.*): *Philosophie Mathématique*, pp. 67–72. Paris: Hermann.

Curry, H. B. [1951]: *Outline of a Formalist Philosophy of Mathematics*. Amsterdam: North-Holland.

Curry, H. B. [1963]: *Foundations of Mathematical Logic*. New York: McGraw-Hill.

Curry, H. B. [1965]: 'The Relation of Logic to Science', in P. Bernays and S. Dockx (*eds.*): *Information and Prediction in Science*, pp. 79–98. New York and London: Academic Press.

Feferman, S. [1968]: 'Autonomous Transfinite Progressions and the Extent of Predicative Mathematics', in B. van Rootselaar and J. F. Staal (*eds.*): *Logic, Methodology and Philosophy of Science III*, pp. 121–35. Amsterdam: North Holland.

Fraenkel, A. A. [1927]: *Zehn Vorlesungen über die Grundlegung der Mengenlehre*. Leipzig and Berlin: B. G. Teubner.

Fraenkel, A. A., Bar Hillel, Y and Levy, A. [1973]: *Foundations of Set Theory*. Second edition. Amsterdam: North Holland.

Gödel, K. [1931]: 'Discussion zur Grundlegung der Mathematik', *Erkenntnis*, **2**, pp. 147–8.

Gödel, K. [1938]: 'The Consistency of the Axiom of Choice and of the Generalized Continuum Hypothesis', *Proceedings of the National Academy of Science, U.S.A.*, **24**, pp. 556–7.

Gödel, K. [1944]: 'Russell's Mathematical Logic', in P. A. Schilpp (*ed.*): *The Philosophy of Bertrand Russell*, pp. 125–53. New York: Tudor. Reprinted in P. Benacerraf and H. Putnam (*eds.*): *Philosophy of Mathematics, Selected Readings*, pp. 211–32.

Gödel, K. [1947]: 'What is Cantor's Continuum Hypothesis?', *American Mathematical Monthly*, **54**, pp. 515–25.

Gödel, K. [1964]: 'What is Cantor's Continuum Hypothesis?', in P. Benacerraf and H. Putnam (*eds*): *Philosophy of Mathematics, Selected Readings*, pp. 258–73. Revised and expanded version of Gödel [1947].

Herbrand, J. [1930]: 'Les Bases de la Logique Hilbertienne', *Revue de la Métaphysique et de la Morale*, **37**, pp. 243–55

Heyting, A. [1967]: 'Weyl on Experimental Testing of Mathematics', in I. Lakatos (*ed.*): [1967], p. 195.

Hilbert, D. [1925]: 'Über das Unendliche', *Mathematische Annalen*, **95**, pp. 161–90. English translation in P. Benacerraf and H. Putnam (*eds.*): *Philosophy of Mathematics, Selected Readings*, pp. 134–51. Oxford: Basil Blackwell.

Kalmár, L. [1959]: 'An Argument against the Plausibility of Church's Thesis', in A. Heyting (*ed.*): *Constructivity in Mathematics*, pp. 72–80. Amsterdam: North-Holland.

KALMÁR, L. [1967]: 'Foundations of Mathematics—Whither Now?', in I. Lakatos (*ed.*), [1967], pp. 187–94.

KEMENY, J. G. [1958]: 'Undecidable Problems of Elementary Number Theory', *Mathematische Annalen*, **135**, pp. 160–9.

KLEENE, S. C. [1943]: 'Recursive Predicates and Quantifiers', *Transactions of the American Mathematical Society*, **53**, pp. 41–73.

KLEENE, S. C. [1952]: *Introduction to Metamathematics*. Amsterdam: North-Holland.

KLEENE, S. C. [1967]: 'Empirical Mathematics?', in I. Lakatos (*ed.*): [1967], pp. 195–6.

KLEENE, S. C. and ROSSER, J. B. [1935]: 'The Inconsistency of Certain Formal Logics', *Annals of Mathematics*, **36**, pp. 630–6.

KREISEL, G. [1967*a*]: 'Informal Rigour and Completeness Proofs', in I. Lakatos (*ed.*): [1967], pp. 138–71.

KREISEL, G. [1967*b*]: 'Reply to Bar-Hillel' in I. Lakatos (*ed.*): [1967], pp. 175–8.

KREISEL, G. [1967*c*]: 'Comment on Mostowski', in I. Lakatos (ed.): [1967], pp. 97–103.

KUHN, T. S. [1963]: 'The Function of Dogma in Scientific Research', in A. C. Crombie (*ed.*): *Scientific Change*, pp. 347–69.

LAKATOS, I. [1962]: 'Infinite Regress and the Foundations of Mathematics', *Aristotelian Society Supplementary Volumes*, **36**, pp. 155–84.

LAKATOS, I. [1967]: 'A Renaissance of Empiricism in the Recent Philosophy of Mathematics?', in I. Lakatos (*ed.*): [1967], pp. 199–202.

LAKATOS, I. (*ed.*) [1967]: *Problems in the Philosophy of Mathematics*. Amsterdam: North-Holland.

LAKATOS, I. [1970]: 'Falsification and the Methodology of Scientific Research Programmes', in I. Lakatos and A. Musgrave (*eds.*): *Criticism and the Growth of Knowledge*. Cambridge University Press. pp. 91–196.

LAKATOS, I. [1971]: 'Popper zum Abgrenzungs- und Inductionsproblem', in H. Lenk (*ed.*): *Neue Aspekte der Wissenschaftstheorie*, pp. 75–110. Braunschweig: Vieweg. In English as 'Popper on Induction and Demarcation', in P.A. Schilpp (*ed.*): *The Philosophy of Karl Popper*, pp. 241–73. LaSalle: Open Court, 1974.

LAKATOS, I. [1977]: 'The Significance of Non-Standard Analysis for the History and Philosophy of Mathematics', *forthcoming*.

LEVY, A. and SOLOVAY, R. M. [1967]: 'Measurable Cardinals and the Continuum Hypothesis', *Israeli Journal of Mathematics*, **5**, pp. 234–48

LUSIN, N. [1935]: 'Sur les Ensembles Analytiques Nuls', *Fundamenta Mathematica*, **25**, pp. 109–31

MARTIN, D. A. and SOLOVAY, R. M. [1970]: 'Internal Cohen Extensions', *Annals of Mathematical Logic*, **2**, pp. 143–78.

MEHLBERG, M. [1962]: 'The Present Situation in the Philosophy of Mathematics', in B. M. Kazemier and D. Vuysje (*eds.*): *Logic and Language: Studies Dedicated to Professor Rudolf Carnap on the Occasion of his Seventieth Birthday*, pp. 69–103. Dordrecht: Reidel.

MOSTOWSKI, A. [1955]: 'The Present State of Investigations on the Foundations of Mathematics', *Rozprawy Matematyczne*, **9**. Compiled in collaboration with A. Grzegorczyk, S. Jaśkowski, J. Loś, S. Mazur, H. Rasiowa, and R. Sikorski.

MYHILL, J. [1960]: 'Some Remarks on the Notion of Proof', *The Journal of Philosophy*, **57**, pp. 461–71.

NEUMANN, J. VON [1947]: 'The Mathematician', in R. B. Heywood (*ed.*): *The Works of the Mind*, pp. 180–96. Chicago: University of Chicago Press.

POPPER, K. R. [1935]: *Logik der Forschung*. Vienna: Springer. Translated into English as *The Logic of Scientific Discovery*. London: Hutchinson, 1959.

QUINE, W. V. O. [1941*a*]: 'Element and Number', *Journal of Symbolic Logic*, **6**, pp. 135–49. Reprinted in *Selected Logical Papers*, 1966, pp. 121–40. New York: Random House.

QUINE, W. V. O. [1941*b*]: 'Review of Rosser: "The Independence of Quine's Axioms *200 and *201" ', *Journal of Symbolic Logic*, **6**, p. 163.

QUINE, W. V. O. [1953]: 'On ω-inconsistency and a So-called Axiom of Infinity', *Journal of Symbolic Logic*, **18**, pp. 119–24. Reprinted in *Selected Logical Papers*, 1966, pp. 114–20. New York: Random House.

QUINE, W. V. O. [1963]: *Set Theory and Its Logic*. Cambridge, Massachusetts: Harvard University Press.

QUINE, W. V. O. [1958]: 'The Philosophical Bearing of Modern Logic', in R. Klibansky (*ed.*): *Philosophy in the Mid-Century*, I. pp. 3–4 Firenze: La Nuova Italia.
QUINE, W. V. O. [1965]: *Elementary Logic*. Revised edition. New York: Harper Torchbooks.
ROSSER, J. B. [1937]: 'Gödel's Theorems for Non-Constructive Logics', *Journal of Symbolic Logic*, **2**, pp. 129–37.
ROSSER, J. B. [1941]: 'The Independence of Quine's Axioms *200 and *201', *Journal of Symbolic Logic*, **6**, pp. 96–7.
ROSSER, J. B. [1953]: *Logic for Mathematicians*. New York: McGraw-Hill.
ROSSER, J. B. and WANG, H. [1950]: 'Non-Standard Models for Formal Languages', *Journal of Symbolic Logic*, **15**, pp. 113–29. (*Errata, ibid.*, p. iv.)
RUSSELL, B. A. W. [1901]: 'The Study of Mathematics' in *Philosophical Essays*. The references are to the reprint in *Mysticism and Logic*, pp. 48–58. London: George Allen and Unwin, 1963.
RUSSELL, B. A. W. [1924]: 'Logical Atomism', in J. H. Muirhead (*ed.*): *Contemporary British Philosophy: Personal Statements, First Series*. pp. 357–83 Reprinted in R. C. Marsh (*ed.*): *Logic and Knowledge*, pp. 323-43. London: George Allen & Unwin, 1956.
RUSSELL, B. A. W. [1959]: *My Philosophical Development*. London: George Allen & Unwin.
RUSSELL, B. A. W. and WHITEHEAD, A. N. [1925]: *Principia Mathematica*, **1**. Second edition. Cambridge: Cambridge University Press.
SHOENFIELD, J. [1971]: 'Measurable Cardinals' in R. O. Gandy and C. E. M. Yates (*eds.*): *Logic Colloquium '69*, pp. 19–49. Amsterdam: North Holland.
SIERPINSKI, W. [1935]: 'Sur une Hypothèse de M. Lusin', *Fundamenta Mathematica*, **25**, pp. 132–5.
SOLOVAY, R. M. and TENNENBAUM, S. [1967]: 'Iterated Cohen Extensions and Souslin's Problem', *Annals of Mathematics*, **94**, pp. 201–45.
SPECKER, E. P. [1953]: 'The Axiom of Choice in Quine's *New Foundations for Mathematical Logic*', *Proceedings of the National Academy of Sciences, U.S.A.*, **39**, pp. 972–5.
TARSKI, A. [1939]: 'On Undecidable Statements in Enlarged Systems of Logic and the Concept of Truth', *Journal of Symbolic Logic*, **4**, pp. 105–12.
TARSKI, A. [1954]: 'Comments on Bernays: "Zur Beurteilung der Situation in der Beweistheoretischen Forschung"', *Revue Internationale de Philosophie*, **8**, pp. 17–21.
TURING, A. M. [1939]: 'Systems of Logic Based on Ordinals', *Proceedings of the London Mathematical Society*, **45**, pp. 161–228.
WEYL, H. [1928]: 'Diskussionsbemerkungen zu dem Zweiten Hilbertschen Vortrag über die Grundlagen der Mathematik', *Abhandlungen aus dem Mathematischen Seminar der Hamburgischen Universität*, **6**, pp. 86–8.
WEYL, H. [1949]: *Philosophy of Mathematics and Natural Science*. Princeton: Princeton University Press.
WHEWELL, W. [1860]: *On the Philosophy of Discovery*. London: Parker.

Brit. J. Phil. Sci. 27 (1976), 225–229 *Printed in Great Britain*

On the Logical Status of Equivalence Principles in General Relativity Theory

by MENDEL SACHS

I should like to put forward the thesis that, in contrast with early statements by Einstein and others on the axiomatic basis of the theory of general relativity, the *principle of equivalence* is not one of its underlying axioms. More fundamental to the mathematical structure of relativity theory is the *principle of correspondence*—which does make use of the various forms of the equivalence principle, under the appropriate approximating circumstances where the formalism asymptotically approaches the classical Newtonian formalism.

1 There are essentially two forms of the *principle of equivalence*. The first, sometimes called 'weak equivalence', asserts that, *in principle*, the inertial mass of a body, m_i, is precisely equal to its gravitational mass, m_g. This assertion is added to the existing logical structure of the dynamical theory of matter in motion because m_i and m_g are conceptually distinct. That is, there is no *a priori* reason why it should necessarily follow that $m_i = m_g$, even though experimental observations confirm this equality to high accuracy.

Recall that 'inertial mass' refers to an intrinsic property of matter in motion. In classical physics, it is the measure of a body's resistance to a change of state of motion, and it is represented mathematically as the constant of proportionality in

$$\mathbf{F} = m\mathbf{a} \tag{1}$$

between the *cause* of a change of a state of motion (the external force **F**) and the acceleration **a**, which is the *effect*.

On the other hand, the gravitational mass, m_g, is the constant of proportionality in

$$W = m_g(GM/R^2) \tag{2}$$

between the gravitational force, W, of one body, on the observed body, causing it to have weight, and the parameters (GM/R^2), where G is Newton's universal gravitational constant, M is the mass of the weight producing body, and R is the distance from m_g to the centre of mass of

Received 29 *May* 1975

M (assuming the latter to be spherically symmetric in this example). Thus, m_i has to do with a property of matter in motion, whether or not the gravitational force, *per se* is causing that motion, and m_g has to do with the action of the gravitational field, in particular, causing this body to have weight.

These remarks refer to classical physics. From the view of general relativity theory, Newton's gravitational force law (2) is not a valid law, in principle. Thus, the gravitational mass parameter, m_g, is not defined, in an exact sense, in general relativity theory—though it appears in the 'local limit' of the theory. But the inertial mass is conceptually defined, *globally*, in general relativity, though Newton's second law of motion (1) is also not valid here, except as a local approximation.[1] It seems to me, then, that since m_g is generally undefined, it is logically false to assert the (weak form of the) *equivalence principle* as a *bona fide underlying axiom* of the theory of general relativity.

The actual underlying principle that this relates to in Einstein's theoretical construction is the *principle of correspondence*—the assertion that under the proper conditions (of sufficiently small relative separations and speeds of interacting matter), the general field formalism must asymptotically approach the Newtonian action-at-a-distance formalism for interacting bodies. Thus, Newton's formalism does provide a useful approximation; but only in this limit can one talk about the equality of inertial and gravitational mass. However, even in this limit, where the Newtonian approximation is used, we are aware that certain terms in Einstein's equations that are neglected because of their smallness, are nevertheless there all the time! And so long as they are there, in principle, m_g is in fact undefined. Thus, the statement $m_i = m_g$ is not at all meaningful in the exact sense of an *underlying* principle, though it must be used in the application of the *principle of correspondence*.

2 The other statement of the principle of equivalence (sometimes called 'strong equivalence') does not directly refer to gravitational mass, but rather refers to an equation of motion of a test body in general relativity—the geodesic equation. The idea here is that there are two equivalent ways of talking about the gravitational action of one body on another: (1) the body is moving freely, though nonuniformly, with a type of motion that would be produced by a gravitational force. Or (2), instead of considering that the body is actually in a state of motion, one may equivalently consider it to be at rest, though acted upon by an external gravitational force.

---

[1] For further critical comments on the conventional interpretation of the equivalence principle, and a mathematical treatment of the role of the 'local approximation', in this context, see Synge [1960], preface, p. ix and ch. III, section 9.

In general relativity theory, this comparison comes from the following argument: According to classical physics, Newton's first law of motion

$$\ddot{x}^k = 0 \qquad (k = 1,2,3) \tag{3}$$

integrates to give the straight line path for the trajectory of a body,

$$x^k = x_0^k + \dot{x}_0^k t, \tag{4}$$

when it is not acted upon by any external force. That is, this is the classical 'geodesic' for a free body. Thus, if a body should be observed to move along a path that is not a straight line, then one must conclude, with Galileo and Newton, that $\ddot{x}^k \neq 0$ (Newton's second law of motion). But this classical geodesic entails the assumption of a Euclidean space. If the space were instead defined with a non-Euclidean geometry, as is the case in general relativity theory, then without introducing any external force, the geodesic equation that replaces (3) is:

$$\ddot{x}^k = -\Gamma^k_{\alpha\beta}\dot{x}^\alpha\dot{x}^\beta \qquad (\alpha,\beta = 1,2,3,4). \tag{5}$$

The affine connection field, $-\Gamma^k_{\alpha\beta}$, depending on the curvature of space-time, multiplied by $\dot{x}^\alpha\dot{x}^\beta$ and summed over α,β then plays the role of $F^k/m \neq 0$ of Newton's theory, though it has a different (more complicated) form. The trajectories of a test body, following from the solutions of equation (5), are curved paths rather than straight lines. Thus, Einstein proposed that the 'free' curved motion of a test body, in a Riemannian space, is in some way equivalent to the effect of some external force acting on this body. Indeed, one of the outstanding successes of Einstein's theory was that the curved paths, so-predicted, exactly fit the observed paths, such as the paths of planetary orbits, *in better agreement with the data* than Newton's theory provided (*e.g.* Einstein's prediction of the perehelion precession of a planetary orbit).

But even this statement of the equivalence principle, in terms of the relation of the geodesic equation to a force law, is an approximation. For while we assume here that the motion of a 'test mass' m is caused by a background gravitational field, represented by the geometrical field $\{\Gamma^k_{\alpha\beta}\}$, the presence of the test body with mass m is assumed here to have no influence on this geometrical field. Though this is a very useful approximation in many applications of general relativity theory, *it is not exact*, in principle! According to the full logical content of the theory of general relativity, the affine connection field, $\{\Gamma^k_{\alpha\beta}\}$, is supposed to be a representation for *all* of the matter of a closed system. If one can (artificially) isolate a part of the system, calling its inertial mass m, the rest of the closed system must then recoil under the influence of m's motion, small, *e.g.*, as the sun's recoil would be due to the earth's motion!

The error made above lies in describing this closed system as a hybrid of a particle (test body) and continuous field (*i.e.* the metrical field of geometry, representing all of the other matter that acts on m). This is the same type of superposition of particle and field that Faraday used in his field theory. But in the latter case, it is valid according to Faraday's definition of the field as a linear superposition of *additive fields*. In Faraday's case, these are potential fields of force, due to a system of charged bodies, to be actualized by a test body when it comes into their domain of influence. On the other hand, the field of general relativity theory represents a *closed system*—not obeying the rule of linear superposition, and not additive in any way in terms of a sum of parts.[1]

General relativity theory is strictly a continuum field theory. What appears as a particle can only be an (approximately localised) field concentration—peaked, though still *continuously distributed.* Of course, it is still (fortunately) very useful to approximate a part of the closed system as a *bona fide* particle, with a precisely localised mass. But my point is that this is not what one has, in fact, according to the theory of general relativity, and therefore the *principle of equivalence* (even in the 'strong form') indeed cannot claim the logical status of an underlying axiom of the theory.

3 With this argument, I conclude that the theory of general relativity, as formulated by Einstein, is based on two underlying axioms, one strong and one weak one. The strong one is the *principle of covariance*—the assertion of the objectivity of the laws of nature with respect to space-time transformations between arbitrarily moving reference frames. The relation between matter and geometry follows from the interpretation of the logic of space-time (*i.e.* its geometry and algebra), as representative of matter that the laws of nature refer to. The variability of the matter content of a closed system then necessitates the variability of the metric relations between the coordinates of space-time—thereby requiring a departure from Euclidean geometry.

The *principle of correspondence* is a weak axiom of the theory since there is no *a priori* reason to maintain this axiom for all time, if it should be possible to construct a new generally covariant theory, equally successful in predicting the data, though without imposing the classical limit in the theory. Nevertheless, a brief look at the history of science convinces me that the *correspondence principle* will remain, and will continue to apply as science progresses in the future.

With the *principle of correspondence*, it follows that of all possible non-Euclidean geometries to choose from according to our present knowledge, the Riemannian geometry is the one to use in general relativity theory. This is because it alone asymptotically approaches the Euclidean system, in the limit of small relative spatial separations and speeds between interacting matter.

---

[1] Sachs [1973], ch. IV.

In my own research programme in general relativity theory,[1] I have added one more axiom to complete the unification of the explanation of the inertial manifestation of matter with its force manifestations. This is what I have called the 'generalised Mach Principle', an assertion that in addition to the inertial mass, all previously considered 'intrinsic' properties of matter are in fact not intrinsic, but rather are features of the dynamical coupling of a closed system of matter. This axiom does away with all remnants of atomicity of matter, yielding a purely continuum field theory for its fundamental description.[2]

State University of New York at Buffalo

REFERENCES

SACHS, M. [1971]: 'A New Theory of Elementary Matter, Part I: Philosophical Approach and General Implication, Part II: Electromagnetic and Inertial Manifestations', *International Journal of Theoretical Physics*, **4**, pp. 433–51, 453–76.

SACHS, M. [1972]: 'A New Theory of Elementary Matter, Part III: A Self-Consistent Field Theory of Electrodynamics and Correspondence with Quantum Mechanics, Part IV: Two-Particle Systems: The Particle-Antiparticle Pair and Hydrogen', *International Journal of Theoretical Physics*, **5**, pp. 35–53, 161–97.

SACHS, M. [1973]: *The Field Concept in Contemporary Science*. Illinois: Thomas Springfield.

SYNGE, J. L. [1960]: *Relativity: The General Theory*, New York: North Holland.

[1] Sachs [1971] and [1972].

[2] I wish to thank the members of the faculty of the Institute of Physics, National Tsing Hua University, Taiwan, China, for their warm hospitality during May 1975, when this note was prepared.

Brit. J. Phil. Sci. 27 (1976), 231–249 *Printed in Great Britain*

The Pragmatist Theory of Truth

by SUSAN HAACK

1 *Introduction.*
2 *Sketch of the Theory.*
3 *Some Replies to Some Critics, and Some New Criticisms.*
  (*a*) *A Definition or a Criterion of Truth?*
  (*b*) *Truth and Utility.*
  (*c*) *Truth and Verifiability.*
  (*d*) *The Pragmatist Theory and the T-Schema.*
  (*e*) *A Subjectivist Theory?*
  (*f*) *Truth as the End of Enquiry.*
4 *Some Concluding Remarks.*

1 INTRODUCTION

How can the mere pragmatist feel any duty to think truly?

(Royce)

My failure in making converts to my conception of truth seems, if I may judge from what I hear, . . . almost complete. An ordinary philosopher would feel disheartened, and a common choleric sinner would curse God and die, after such a reception. But instead of taking counsel of despair, I make bold to vary my statements, in the faint hope that repeated drippings may wear upon the stone.

(James, 'A Word More About Truth' in his [1909], p. 136)

James's pessimism was amply justified; the unpopularity of the pragmatist theory of truth has persisted, at least on this side of the Atlantic, to the present day. The following passage is typical:

> William James, the originator of the [pragmatist] theory [of truth], took over the central idea from C. S. Peirce (*sic*), but altered it in the process. Peirce had put forward practical usefulness as a criterion of meaningfulness . . . James applied this idea (perhaps confusedly) to truth in the attempt to provide a down-to-earth substitute for certainty within the theory of knowledge. But merely to reject the search for certainty by putting something less in its place without diagnosis of the reasons for the demand for certainty in the first place is to some extent an abrogation of the philosopher's responsibility. The pragmatic theory cannot therefore be put on the same level as the . . . correspondence and coherence theories. I shall not discuss it further. (It would not be unfair to say that it is founded on a muddle.)
>
> (Hamlyn [1970], p. 119)

Received 28 *August* 1975

Every sentence of this passage, with the single exception of 'I shall not discuss it further', is either false or seriously misleading. Peirce, not James, originated the pragmatist theory of truth; Peirce's theory of meaning does not equate significance with practical usefulness, and neither does James's theory of truth straightforwardly equate truth with practical usefulness; Dewey provided in his [1929] a particularly acute diagnosis of the classical quest for certainty; and the pragmatist theory of truth contains substantial coherence and correspondence elements.

It would, I fear, be generous to describe Hamlyn's account as founded on a muddle; and I shall not discuss it, at least directly, further.

What I hope to do, though, is to convince you, by showing that the most influential criticisms have been based on a very inadequate understanding of the theory, that this kind of dismissive attitude is unjustified. The pragmatist theory of truth is not, to be sure, without difficulties; but there is, I think, a good deal to be learned from it, both from its strengths and from its weaknesses. It is, at any rate, of sufficient interest amply to repay the effort of setting the historical record straight.

2 SKETCH OF THE THEORY

By 'the pragmatist theory of truth' I shall understand a set of interlocking theses, to be found in the works of Peirce, Dewey and James, which may together be regarded as constituting a theory of truth. (I shall not consider the views of Schiller.) Not all these theses will be found in all three writers, but most of them will be found in at least two, though with varying emphasis. There are dangers, of course, in speaking of '*the* pragmatist theory of truth' at all, for there are interesting and important differences between the pragmatists on these issues. But I hope that a joint treatment will enable me to establish, first, that the theory originally offered by Peirce and subsequently adopted by Dewey was considerably extended by James, that it was upon James's version of the theory that the most influential criticisms fell, and that many of these criticisms can be seen to be misguided once James's views are placed in the context of the underlying Peircean theory; and second, that some of the differences between Peirce's and James's versions can be seen as resulting from their different reactions to an internal tension in the common part of their theory. I shall restrict myself to consideration of the truth of straightforwardly factual beliefs; the application of the theory to, for example, mathematical or metaphysical beliefs would require a detailed account of the pragmatists' views on mathematics and metaphysics.

For Peirce, truth is the end of inquiry. Here 'end' is ambiguous, between 'aim' and 'final state'. This ambiguity gives rise to some difficulty, but for the present I leave it as it stands. Peirce's theory of inquiry goes, in outline, as follows: a person in a state of (real, not Cartesian) doubt, struggles to attain instead a state of fixed belief. Peirce argues that some methods of acquiring beliefs—the method of tenacity, the method of authority, and the *a priori* method—are unsatisfactory because they are inherently unstable. A person using one of these methods will acquire an opinion, but different people will thereby acquire different opinions, and the existence of rival opinions will raise doubt all over again. Only one method, the Scientific Method, is stable; enables one, that is, to acquire a belief that will not be shaken.

The Scientific Method, alone among methods of acquisition of beliefs, has this virtue, because it is constrained by Reality, which is independent of our beliefs about it. Beliefs acquired by the use of the Scientific Method are caused by Real Things; so the use of the Scientific Method cannot but lead, eventually, to a stable consensus.

Since inquiry is prompted by doubt, and ended only with the acquisition of a stable belief, and since the Truth is that stable consensus which the Scientific Method will eventually achieve, it follows that the true is, in a certain sense, satisfactory to believe; satisfactory because stable. (Since Peirce's theory of belief is behaviourist, this satisfactoriness is, in a way, 'practical'.)

James pays little attention to the theory of inquiry. He agrees that truth is correspondence with reality, but, even more vehemently than Peirce, insists on knowing what difference a belief's correspondence with reality might make. His major contribution is thus a substantial extension of the third thesis, that the truth is satisfactory to believe. The benefit of holding true beliefs, according to James, is that if what one believes is true one is, so to speak, guaranteed against recalcitrance on the part of experience. No doubt one could get along perfectly well, for a bit, holding false beliefs; but, James thinks, one would eventually be caught out.

This provides a sensible interpretation of those, perhaps, incautious remarks of James's which Moore found 'silly', that:

> The true is the name of whatever proves itself to be good in the way of belief . . .
> ('What Pragmatism Means', in [1907], p. 59)

and

> The true . . . is only the expedient in our way of thinking.
> ('Pragmatism's Conception of Truth' in [1907], p. 145)

By beliefs which are 'good' or 'expedient' or which 'pay' James means beliefs which are safe from the danger of inconsistency with subsequent experience. That this is a fair interpretation becomes clear from the following passages, the first appearing shortly after the identification of the true with the good to believe, the second immediately after the identification of the true with the expedient to believe:

> . . . what is better for us to believe is true *unless the belief incidentally clashes with some other vital benefit*. Now in real life what vital benefits is any particular belief most liable to clash with? What indeed except the vital benefits yielded by other benefits when these prove incompatible with the first ones?

> . . . expedient in the long run and on the whole of course; for what meets expediently all the experience in sight won't necessarily meet all farther experience equally satisfactorily. Experience, as we know, has ways of *boiling over*, and making us correct our present formulas.

So, for James as for Peirce, the true is satisfactory (useful, expedient, good) to believe, because it is safe from overthrow by subsequent experience. James, however, amplifies this thesis by an account of the way one modifies, in the face of inconsistency with a new experience, the beliefs one previously held true; one aims to maximise the conservation of the old belief set while restoring consistency. The likeness to Quine's epistemology ([1951]) is striking.

True beliefs, James frequently comments, are those which are verifiable. By this he means that those beliefs are true which, in the long run, are corroborated or confirmed by experience. (He does not distinguish corroboration from confirmation, as subsequent writers have done.) Most of the beliefs we take to be true are, James admits, actually verified at best only very indirectly. Our beliefs are like banknotes, they 'pass' so long as no-one challenges them; but, once again like the financial system, the system of beliefs would collapse were it not for actual direct verifications at some points.

I have so far stressed the similarities between Peirce's and James's view. But there is a difference of emphasis which it is important to consider. Peirce is preoccupied with the Truth, that is, the totality of individual truths. James, by contrast, is primarily interested in the individual truths, finding the Truth with a capital T a somewhat spectral and uninteresting abstraction. He is aware that one *could* say that some propositions just are true (or false) even though no-one has ever verified (or falsified) them or, indeed, even entertained them; but he finds this way of talking relatively *pointless*. (This view bears some resemblance to Dummett's rather stronger, Intuitionist view of truth in his [1959].) In fact James quite often insists that the Truth is growing corpus, that is, that new truths come into existence as human knowledge grows. This emphasis on the growth of truth can be understood as a consequence of James's switch of Peirce's emphasis on ***the totality of truths in the long run*** to an emphasis on ***individual truths in the short run***. This difference of emphasis can probably be traced in its turn to the contrast between Peirce 's realism and James's nominalism: Peirce did not share James's antipathy for abstractions. His nominalism also underlies James's tendency to prefer to speak of actual verifications, where possible, rather than verifiability (unactualised possible verifications being somewhat embarassing to a nominalist). This tendency, as I shall argue subsequently, gets him into a serious difficulty.

The same difference of emphasis is reflected in another feature of James's presentation. At any particular time, James argues (short of the fictional Long Run, that is) the evidence available to us may be insufficient to decide between competing beliefs; and then our choice *will* be a matter of taste:

> ... sometimes alternative theoretic formulas are equally compatible with all the truths we know, and then we choose between them for subjective reasons ... we follow 'elegance' or 'economy'.
>
> ('Pragmatism's Conception of Truth', in his [1907], p. 142)

Indeed, in at least one place ('What Pragmatism Means', in his [1907] p. 44) James refers to the possibility that, even after *all* the data are in, alternative theories, between which we should choose on such 'aesthetic' grounds as simplicity, economy *etc.* may remain. Although the resemblance to Quine's epistemology is, once again, striking, two points of contrast should be noted. First, Quine takes it that there could be a real difference between two such theories, whereas James, I think, would rate the difference merely verbal; second, Quine would not, I think, so readily admit that considerations such as simplicity and economy are *purely* subjective. Both points will be relevant in later parts of the paper.

Dewey follows Peirce in stressing that truth is the end of inquiry, though he adds considerably to the theory of inquiry. In *Logic, the Theory of Inquiry* ([1938], p. 345 n.) he simply quotes Peirce's as 'the best definition of truth'; Dewey prefers to replace the word 'truth' by 'warranted assertibility' to emphasise that the truth is precisely what the method of inquiry warrants us in asserting.

Dewey contributes to the theory an important insight into the role of truth. Truth, or warranted assertibility, characterises those beliefs to which we give the honorific title, 'knowledge'. It has subsequently been commonplace (though not, of course, uncontroversial) to analyse 'x knows that p' along the lines of 'x believes p, p is true and x has good reasons for his belief that p'. Dewey, interestingly enough, *merges* the truth with the warrant requirement.

The central thesis of the theory may be summarised thus:

Truth—
is the end of inquiry
is correspondence with reality } Peirce
is satisfactory to believe
is coherence with experience—verifiability } James } Dewey
[is a growing corpus]
entitles belief to be called 'knowledge'.

3 SOME REPLIES TO SOME CRITICS, AND SOME NEW CRITICISMS

(*a*) *A definition or a criterion of truth?*

Some early critics suggested that the pragmatists had confusedly presented a criterion of truth as though it were a definition of truth; that their theory provided at best, a test of truth, but was offered as if it gave an analysis of the meaning of 'true':

> The test of truth and the meaning of truth are . . . completely identified
> (Pratt, What is Pragmatism? [1909], p. 89)

> . . . if pragmatists only affirmed that utility is a *criterion* of truth, there would be much less to be said against their view . . . The arguments of the pragmatists are almost wholly directed to proving that utility is a *criterion*; that utility is the *meaning* of truth is then supposed to follow.
> (Russell, 'James's Conception of Truth', [1908], p. 121)

Interestingly enough, the same distinction is used by Rescher in *The Coherence Theory of Truth* ([1973]), where he tries to give, as he thinks, a more sympathetic account of the pragmatist theory by presenting it as precisely, a criterion, but not a definition, of truth.

But these criticisms, and Rescher's attempted rehabilitation, are both inappropriate. For the pragmatists' view of meaning is such that a dichotomy between definitions and criteria would have been entirely unacceptable to them. It is a fundamental tenet of pragmatism (Peirce sometimes says, *the* fundamental tenet) that meaning is given by reference to experiential consequences. Thus, Peirce:

> . . . consider what effects, which might conceivably have practical bearings, we conceive the object of our conception to have. Then, our conception of those effects is the whole of our conception of the object.
>
> . . . let us ask what we mean by calling a thing *hard*. Evidently that it will not be scratched by many other substances. The whole conception of this quality, as of every other, lies in its conceived effects.
> ('How to make our ideas clear', [1878], p. 124)

and James:

> There can *be* no difference that *makes* no difference
> ('What Pragmatism means', in [1907], p. 45)

This view of meaning raises many interesting issues (its relation to Logical Positivism and Operationalism, for instance, and its role in persuading C. I. Lewis of the need for an implication relation stronger than the material conditional); but for present purposes it is sufficient to notice that it certainly does not allow a distinction between what 'true' means, and what difference it would make, whether a sentence were true or false. The pragmatists hoped to explain what 'true' means precisely by investigating what difference it makes whether one's beliefs are true or false. As James puts it:

> Pragmatism . . . asks its usual question. "Grant an idea or belief to be true", it says, "what concrete difference will its being true make in any one's actual life? How will the truth be realised? What experiences will be different from those which would obtain if the belief were false? What, in short, is the belief's cash-value in experimental terms?"
>
> ('Pragmatism's Conception of Truth', in [1907], p. 133)

So the criticism that the pragmatists 'confuse' definition and criteria is totally inappropriate, since their theory of meaning quite deliberately equates the two; if this *is* a confusion, it must be shown, by a critique of their theory of meaning, *why* it is. I shall argue, later, however, that Peirce's equation of definitions and criteria does lead to a difficulty because of his fallibilist epistemology.

(*b*) *Truth and utility*[1]

Both Moore and Russell find James's equation of the true with the useful thoroughly unacceptable. They both assume that 'useful' has its usual, everyday sense, and proceed to argue, on the one hand, that some true beliefs can fail to be useful, and on the other, that some useful beliefs can fail to be true. It is notable that Moore and Russell are both assuming, in making this kind of criticism, some other, presumably plain correspondence, theory of truth.

> Is it not clear that we do actually sometimes have true ideas, at times when they are not useful, but positively in the way?
>
> (Moore, 'William James' "Pragmatism" ', [1908], p. 110)

> It seems perfectly possible to suppose that the hypothesis that [other people] exist will always work, even if they do not in fact exist. It is plain . . . that it makes for happiness to believe that they exist . . . But if I am troubled by solipsism, the discovery that a belief in the existence of others is 'true' in the pragmatists' sense is not enough to allay my sense of loneliness.
>
> (Russell, 'James's Conception of Truth', [1908], p. 122)

I have argued in section 2 that when James says that true beliefs are

---

[1] *Cf.* Perkins [1952] and Hertz [1971].

'useful' he should be understood to mean that they are so in the sense of being guaranteed against overthrow by subsequent experience. Neither Moore's nor Russell's criticisms are to the point if this *is* what James meant. And there are—besides the passages I have referred to as confirming my interpretation—also passages where James repudiates the thesis which Moore and Russell attribute to him. He observes, for example, that a belief in the Absolute would afford him a kind of 'moral holiday', that is, that it would be good, in the sense of congenial, to believe; but he rejects this belief, nevertheless, on the grounds that it would be inconsistent with 'other [*sic*] truths'. ('What Pragmatism means' in [1907], p. 61). And compare this comment:

> Above all we find *consistency* satisfactory.
> ('The Pragmatist Account of Truth and its misunderstanders', in [1909], p.192)

James perhaps comes closest to the view which Moore and Russell attribute to him in his discussions of religious belief, about which he is apt to say that that belief is true which best succeeds in making coherent both one's experiences and one's values. And even here, clearly, he is by no means straightforwardly maintaining that truth is a matter of taste or (in the ordinary sense) expediency; rather, he is extending his coherence view to moral as well as empirical beliefs.

The pragmatists' view of the truth as the satisfactory to believe has deep roots in Peirce's epistemology. Peirce's theory of inquiry rests, as I have reported, on the idea that the state of doubt, of not knowing, that is, whether p, is uncomfortable, that it prompts inquiry, and that the discomfort of doubt is ended by the acquisition of a stable belief. This idea in its turn is supported by definitions, due in essentials to Alexander Bain, of belief as a habit of action and of doubt as the interruption of such a habit by novel stimuli. I cannot enter, here, into the question of the adequacy or otherwise of these underlying views, except to comment that Peirce's theory of doubt does not seem to take much account of the fact that doubt or ignorance on some issues may occasion one much less discomfort than doubt or ignorance, on others with which one is—to put it, I fear, question beggingly as well as vaguely—concerned. Ironically, since the pragmatists are so often accused of excessive attention to the merely and vulgarly practical, this comment suggests that Peirce's theory may pay rather less attention to the practical than it should. And this difficulty suggests another: that the theory may be hard put to it to handle the question of the truth-value of propositions which have never been entertained. This, too, will turn out to be significant.

(*c*) *Truth and verifiability*

Since, as I have argued, they did not appreciate the relation between utility, in James's sense, and coherence with the totality of experience, it is not surprising that Moore and Russell also failed to appreciate the close connection between James's claim that the true is the useful, and his claim that the true is the verifiable. But, though they did not appreciate the connection between the two views, they objected to the identification of truth with verification nearly as vehemently as to the identification of truth with utility, and on similar grounds; a belief could be true but never verified. Now there is a view in James which is not straightforwardly vulnerable to this objection: the view, that is, that a belief is true just in case it is *verifiable*. If James is equating truth with verifiability, not with verification, the objection that there are true but unverified beliefs is irrelevant. And there are numerous passages where James concedes that plenty of true, that is to say *verifiable* beliefs, have not yet been verified (*e.g.* 'Pragmatism's Conception of Truth', in [1907] pp. 136–7). Of course, there will still be room for argument about the identification of truth with verifiability, *à propos*, for example, undecidable mathematical or quantum mechanical sentences—argument which will not be made easier by James's rather vague gloss on 'verifiable': if *p* were to be tested it would, eventually, be verified; but at least this thesis does not fall to the simple objections of Moore and Russell. (There is, once again, a similarity to Dummett's views.)

However the matter is considerably more complicated than has yet appeared. As I reported in section 2, James's nominalistic embarrassment about the notion of possible verification in the long run leads him to try to replace 'verifiable' by 'verified' whenever he can, and, in consequence, to maintain that the Truth is a growing corpus, which individual truths join as they are verified. But if a belief is true just in case it is *verifiable*, all these individual truths are true *before* they are verified (which is what James plainly says in his [1909] p. 165); so the Truth consists always of the same truths, and does not, after all, grow. This is, I think, a simple inconsistency in James's view; that is why, on page 236, I relegated the thesis that the Truth is a growing corpus to brackets.

(*d*) *The pragmatist theory and the T-schema*

It seems worth observing that the admission of the thesis that the Truth grows—against which I have been arguing above—would make the pragmatist theory inconsistent with Tarski's material adequacy condition for theories of truth. Tarski requires (in [1931]) that any acceptable theory should entail all instances of

(T) S is true iff p

(where 'S' names p).

In what follows I shall consider only the weaker condition, that any acceptable definition of truth should be consistent with the truth of all instances of (T); for it is in this form, rather than the stronger condition that all instances of (T) be deducible, that Tarski's material adequacy requirement can conveniently be applied to informal truth-definitions. It is not certain that this material adequacy requirement rules out what one might not too unfairly call 'bizarre' theories of truth – such as, say, 'true = asserted in the Bible'; since presumably a serious advocate of such a theory would maintain that *e.g.*

'Warsaw was bombed in 1940' is true iff Warsaw was bombed in 1940 holds, since if he agrees that the lefthand side is false, he will also maintain, if he is wise, that the righthand side is false too. But what the material adequacy condition *does* seem to rule out are theories which are not bivalent, which allow that some sentences are neither true nor false. Thus, if one supposed, as Łukasiewicz did (though I do not) that 'There will be a sea-battle tomorrow' is neither true nor false in advance of the outcome, then

> 'There will be a sea-battle tomorrow' is true iff there will be a sea-battle tomorrow

is presumably not true, since its lefthand side would be false and its righthand side neither true nor false.

One could, indeed, make the rejection of bivalence consistent with the acceptance of the T-schema if one envisaged the adoption of a non-bivalent metalanguage as well as a non-bivalent object language. Thus, if ' 'p' is true' is neither true nor false if 'p' is neither true or false, ' 'p' is true iff p' could be true even though 'p' was truthvalueless. But to motivate this proposal some argument would be required why ' 'p' is true' should be neither true nor false if 'p' is neither true nor false. It seems doubtful whether, in the present case, one could find any very plausible argument to the desired effect. For that certain quantum mechanical sentences are neither verifiable nor falsifiable can itself be verified; so it would seem proper, if one identified 'true' with 'verifiable' at the metalinguistic level also, to admit that that such sentences are true, is false.

Now if one took the view which James sometimes maintains, that a belief becomes true when it is verified, then since at any time there will be beliefs not yet verified or falsified, those beliefs will be, at that time, neither true nor false.

That version which replaces 'verified' by 'verifiable', which, as I have argued, is to be preferred, may or may not be bivalent, depending upon whether or not it is so interpreted as to allow the possibility that some meaningful sentences may be neither verifiable nor falsifiable, and hence, neither true nor false. Some sentences are *practically* incapable of verification or falsification; but it would be possible to maintain that such sentences are nevertheless verifiable or falsifiable, on the grounds that if they *were to be* tested they *would* be verified or falsified. Other sentences are *theoretically* incapable of verification or falsification; for instance certain sentences of quantum mechanics are such that it follows from the theory that they are neither verifiable nor falsifiable. It would be harder, though not perhaps quite impossible, to maintain that such sentences are nevertheless in some weaker ('logical') sense verifiable or falsifiable; but it remains possible to retain bivalence by denying the meaningfulness of these sentences.

Since James is rather unspecific about exactly how he understands 'verifiable' it is not easy to say what view he would have adopted on these questions. However, his theory of meaning would presumably rule that whatever sentences he counted as unverifiable he would also count as meaningless. And only if some meaningful sentences are allowed to be neither true not false does the identification of truth with verifiability threaten to be inconsistent with Tarski's material adequacy requirement.

(*e*) *A subjectivist theory?*

Moore's and Russell's fears that pragmatism would make truth a mere matter of taste are, as I have argued, unfounded. However, the questions whether, and if so in what sense, the pragmatist theory is subjectivist, require further attention.

It is clear, to begin with, that all the pragmatists agree that truth is correspondence with reality. Even James makes the point quite explicitly:

> [Truth] means [our ideas'] 'agreement' with reality'. Pragmatists and intellectualists both accept this definition as a matter of course.
>
> ('Pragmatism's Conception of Truth', in [1907], p. 133)

So far from denying the correspondence theory which their opponents championed, the pragmatists incorporate it as a part of their theory. But they are better aware than their opponents how inadequate, because unspecific, the formula of 'correspondence with reality' is as it stands.

Peirce frequently stresses that Reality is independent of human beliefs about it. Nonetheless, he manifests some embarrassment with his notion of Reality, since, as he is well aware, he can not prove that an external and independent Reality exists. As he puts it: since he uses the idea of Reality as the foundation for his theory of inquiry, he cannot use that theory to show that there is such a thing as Reality. However, he offers, in support of the hypothesis of Reality the following, rather less than conclusive, considerations:

(i) Inquiry leads to no doubt of the existence of such a Reality: so the theory and its under-pinning are in harmony.
(ii) Doubt arises when one is undecided between a belief and its contradictory, which suggests that there is some *one* thing to which belief should conform.
(iii) The extraordinary success of the Scientific Method in leading to consensus can hardly be an accident, and would be explicable on the hypothesis of Reality.

(The last of these arguments has an independent interest, since it seems to support the hypothesis that Peirce took his theory of Scientific Method actually to describe the methods of practising scientists.) There are, though, some passages where Peirce apparently turns his theory on its head:

> . . . as what anything really is, is what it may finally come to be known to be in the ideal state of complete information . . . reality depends on the ultimate decision of the community . . .
>
> ('Some Consequences of Four Incapacities', [1868], p. 72)

Although in 'The Fixation of Belief' [1877] he sounds confident about the view that agreement depends on Reality, not *vice-versa*, in 'How to Make Our Ideas Clear', [1878] he comments that

> The opinion which is fated to be ultimately agreed by all who investigate is what we mean by the truth, and the object represented by this opinion is the real. *That is the way I would explain reality.*
>
> (p. 133, my italics)

James is less preoccupied with establishing the independence of reality, and more preoccupied with stressing that reality is experientially accessible:

> The only *real* guarantee we have against licentious thinking is the circumpressure of reality itself, which gets us sick of concrete errors, whether there be a trans-empirical reality or not.
>
> ('Humanism and Truth', in his [1909], p. 72)

At this point it becomes essential to look, at least briefly, at the development of Peirce's view of reality. Although Peirce was an ontological realist at least since 1871, in his earlier work he did not accept epistemological realism, but subscribed, instead, to a kind of phenomenalism. His ontological realism allows a reality independent of our thought about it, which, as a constraint on Scientific Method, guarantees the eventual agreement of which it is the object. But his phenomenalism regards reality as merely a construction devised to order phenomena, which, since Peirce's doctrine that there are no incognisables ruled out appeal to possible but nonactual phenomena, could not allow there to be an 'object' of the final opinion unless and until that final opinion is reached. In the 1870s Peirce tried to resolve this conflict by appeal to the infinite future; by the 1890s he had resolved it by rejecting phenomenalism (see Murphey [1961], especially pp. 169, 376–7, O'Connor [1964], and Haack [1975], for more detailed discussion of this development). After about 1890, then, Peirce combined ontological with epistemological realism, a doctrine of direct perception. James's pragmatism, however, was always nominalistic. It will be apparent by now that I should disagree with Scheffler's suggestion that the way James goes beyond Peirce is in applying the pragmatic maxim to truth as well as meaning; I should locate an important difference between Peirce and James, rather, in their account of reality (*cf.* Scheffler [1974], Howard [1975]). This, I think, supplies a better explanation of James's greater vulnerability to the charge of subjectivism.

Dewey's views have affinities with Peirce's later position. Dewey observes that pragmatism's opponents insist that truth is correspondence with reality, but then make that reality so remote and inaccessible that it becomes inexplicable how we should ever have the slightest reason to suppose that our beliefs correspond to it. A proper emphasis on the experiential character of reality will, Dewey thinks, serve to banish this mystery; correspondence with reality *is* coherence with the totality of experience.

The pragmatist theory of truth is certainly *not* subjectivist in the sense of identifying truth-for-me with whatever-I-happen-to-find-congenial-to-believe. The satisfactoriness of true beliefs consists in their correspondence with reality. But a further question then arises: how objective, on the pragmatist view, is reality?

It is here that an important tension in the theory begins to come to the surface. Peirce wishes to emphasise the externality and the independence of reality, for the sake of the objectivity of truth; but, though he would wish to hold that reality is independent of what anyone or everyone believes, he has difficulty in going beyond the weaker thesis that reality is independent of what any individual believes, but not of what the Scientific Community, as a whole and in the long run, believes. This tension arises because Peirce's theory of meaning indentifies the sense of an expression with the criterion of its application, the meaning of 'true' in particular, with the criterion of truth. But Peirce doubts that we have certain ways of acquiring knowledge, or of knowing, if we do reach the truth, that we have done so; he is, in short, a fallibilist. This is why he feels the need to appeal to an independent—but, unfortunately, consequently inaccessible—reality, to close the gap which his fallibilism allows to open between the meaning of, and the criterion for, truth.

(Though Popper's account of the pragmatists' views leave a good deal to be desired, some remarks in his [1960] suggest that he has noticed this tension.) This tension in Peirce also manifests itself in his shifts between the thesis that the truth is what the Scientific Community *will* in the long run agree on, and the thesis that the truth is what the Scientific Community *would* in the long run agree upon, if it *did* agree. These shifts will be examined in more detail in the next section. James, interestingly enough, reacts to this tension in a different way: he shuns the appeal to an independent but inaccessible reality, but then finds himself in a state of chronic embarrassment about those truths, which he cannot quite bring himself to deny to be truths, which, given the fallibility of our means of acquiring knowledge, may never be verified or falsified.

(*f*) *Truth as the end of Inquiry*

The thesis that the true is the satisfactory to believe has borne the brunt of criticism from opponents of pragmatism. The first thesis—truth as the end of inquiry—has received relatively less attention. In an article devoted to Dewey's views, however, Russell offers some counter-arguments. Russell points out that, on one interpretation, the thesis apparently entails that whatever beliefs are held by the last man on earth, are true. He comments, rather acidly:

> As [the last man on earth] will presumably be entirely occupied in keeping warm and getting nourishment, it is doubtful whether his opinions will be any wiser than ours.
>
> ('Dewey's New Logic', [1939], p. 145)

But the thesis may be more charitably interpreted, if it is understood to say that the truth is that opinion to which scientific inquiry tends as a limit, then the fact that science might come to an end before this terminal consensus is reached, is irrelevant.

This interpretation, however, is not without difficulties of its own. For it could still be questioned whether science manifests, as the thesis now seems to require, a tendency towards consensus. Peirce, it seems from at least some passages, *does* believe in such a tendency:

> There is a general *drift* in the history of human thought which will lead to one general agreement, one catholic consent.
>
> ([1931–58], IX, § 12)

Russell takes him severely to task for this faith:

> Is this an empirical generalisation from the history of research? Or is it an optimistic belief in the perfectibility of man? Does it contain any element of prophecy, or is it a merely hypothetical statement of what would happen if men

> of science grew continually cleverer? Whatever interpretation we adopt, we seem committed to some very rash assertion.
>
> ('Dewey's New Logic', [1939], p. 146)

So it looks as if the thesis that the Truth is that consensus which use of the Scientific Method would if it continued sufficiently long, reach, may fail for want of an argument that there *is* any such end.

But this objection misses something. It has so far been left vague what, exactly, one is to understand by the 'Scientific Method'. The kind of objection just outlined takes for granted a rough-and-ready understanding in terms, presumably, of those methods which those we call 'scientists' use. Peirce, however, has a theory about Scientific Method, a theory which perhaps offers some explanation of his—as Russell supposed, quite unwarranted—optimism.

According to Peirce, the Scientific Method includes three types of argument: deductive, inductive and abductive. (Abduction Peirce defines as 'studying facts and devising a theory to explain them'.) As so often, Peirce anticipates more recent philosophers of science, for his theory of scientific method strongly resembles the hypothetico-deductive.

What is important for present purposes is, specifically, Peirce's view of induction. Peirce thinks of inductive arguments as, roughly, those which extrapolate probabilities from given data. His view of probability is frequentist; the probability of B given A is given in terms of the proportion of A's which are B's, and, when the A series is infinite, the probability is the limiting frequency, if any. Peirce offers a justification depending upon the fact that induction, as he defines it, is a self-correcting process. Once again, the argument anticipates a better known later version, Reichenbach's 'pragmatic justification' of the straight rule of induction.

Peirce's optimism about the eventual success of the Scientific Method is now more explicable. An essential element in Scientific Method is induction, and induction, Peirce thinks, is such as to yield the true probability, if there is one, eventually. This is not, of course, to say that Peirce's optimism is warranted; to show *that* one would have to devote considerably more detailed attention to his theory of induction than I have time to do. I claim only to have made his optimism explicable.

This suggestion leaves a question open: the question, whether Peirce's theory of scientific method is intended descriptively or prescriptively—as a description of the methods scientists do use, or as a prescription of the methods they should use. My impression is that Peirce is not wholly clear on this point. When he sounds confident that science progresses, it is perhaps because he is taking it to be the case that scientists do, in fact, use what he regards as the Scientific Method; when he seems less sure, it is perhaps because he is thinking of the theory of scientific method as prescriptive only. And sometimes he does sound much less confident than in the last passage quoted. For instance, in a letter of 1908 to Lady Welby he comments:

> I do not say that it is infallibly true that there is any belief to which a person would come if he were to carry his inquires far enough. I only say that that alone is what I call Truth. I cannot infallibly know that there *is* any truth.
> (p. 398)

Peirce's ambiguity on this point disguises a difficulty: if his theory of Scientific Method is intended prescriptively, he is not entitled to appeal to the alleged success of practising scientists in achieving consensus, as he does in his third argument for reality; if, on the other hand, the theory is intended descriptively, he is after all committed to some rather strong thesis about the progress of science, and so is vulnerable to some of Russell's criticisms.

Quine puts an objection which somewhat resembles Russell's. In *Word and Object* ([1960]) he objects to the proposed identification of the Truth with 'the ideal result of applying scientific method outright to the whole future totality of surface irritations', that there is no reason to suppose that there is any such unique result. Quine's objection is based, however, on his own theory of scientific method, which stresses that alternative theories, incompatible with each other, may be compatible with the totality of possible evidence. It is not certain that Peirce's theory of scientific method allows this possibility; and if not Quine's objection is not directly relevant to Peirce's position. James, on the other hand, does seem to envisage such a possibility; but, taking seriously the principle that there can *be* no difference that *makes* no difference, he concludes that two such 'alternative' theories would not *really*, but, presumably, only *verbally* differ from each other. So his view too would avoid Quine's objection.

Both Russell's objection and Quine's objection bear on the supposed uniqueness of the end of inquiry. Even if Peirce's optimism that the Scientific Method will or would eventually yield a consensus can be justified by appeal to his theory of induction, however, there would remain another difficulty, which, curiously, Russell and Quine ignore. If truth is the end of inquiry, then, not only must all beliefs warranted, in the long run, by the Scientific Method, be true, but also, all truths must be, in the long run, warranted by the Scientific Method. But what reason is there to think that the Scientific Method would eventually yield all truths? Is it not likely, on the contrary, that some true propositions will never even be entertained? (On this point *cf.* Ayer [1968]). This difficulty relates to another feature of the pragmatists' views. The subject of 'true' in their writings is, usually, beliefs; and since they stress the importance of community, they clearly intend that different persons should be able to share the same belief. What is not clear, however, is whether 'belief' could be extended to cover propositions which have never been entertained, or whether, if it could not, the pragmatists would be willing to accept the consequence, that only propositions at some time entertained could be true or false. James, whom one might have expected to have been the most willing to admit this, in fact denies it:

> countless opinions 'fit' realities, and countless truths are valid, though no thinker ever thinks them.

But he finds such 'truths' of no consequence:

> . . . all discarnate truth is static, impotent and relatively spectral, full truth being the truth that energises and does battle.
> ('The Pragmatist Account of Truth and its Misunderstanders' in [1909], p. 204)

> . . . the truth with no-one thinking it, is like the coat that fits tho no one has ever tried it on . . .
>
> (p. 205)

This ambivalent attitude exactly parallels, of course, his attitude to 'truths' as yet unverified.

4 SOME CONCLUDING REMARKS

I shall end with some brief, but, I fear, rather vague, comments about what seem to me to be major strengths, and the major weaknesses, of the theory.

The theory is a cosmopolitan one, in that it includes substantial coherence and correspondence elements; and it thereby acquires some of the strengths of the coherence and correspondence theories while avoiding some of the weaknesses. Ironically, in view of the fact that the pragmatist theory has been available for a long time but has never enjoyed much popularity, Quinton commented as recently as 1966 that the direction from which progress in the theory of truth is to be expected is in 'the close interweaving of the coherence and correspondence theories' ('The Foundations of Knowledge', [1966], p. 86). Furthermore, the pragmatist theory avoids that divorce of the theory of truth from epistemology which is apt to make classical correspondence theories unsatisfying. By insisting that one ask what difference the truth or falsity of a belief would make, the pragmatists ensure that their theory of truth connects closely with their theory of knowledge. Russell, to be sure, found their stress on the experiential cash-value of true beliefs distasteful; he was apt to say that Pragmatism was an 'engineers' philosophy', bound to lead to cosmic impiety, or at any rate to fascism. But this stress on the experiential cash-value of true beliefs can do us the important service of raising the neglected question, what, exactly, one should expect of a theory of truth.

Of course, the close connections between the pragmatist theory of truth and their theory of knowledge and theory of meaning mean that the former theory is vulnerable to criticisms directed, in the first instance, at the latter. The acceptability of the thesis that truth is the end of inquiry depends, for instance, on the thesis that the scientific method leads to consensus, and that in turn on a theory of induction, which may involve difficulties which would threaten the whole superstructure. Or again, objections to the pragmatist theory of meaning would be bound to involve consequent difficulties for the theory of truth. And, as I have suggested above, there are indeed tensions between Peirce's pragmatist theory of truth and his fallibilist epistemology. But it is to be hoped that recognition of such tensions may be useful in illuminating some real and important problems which a dogmatic separation of theory of truth and theory of knowledge is apt to disguise.

James once commented that theories generally run through three stages: first, the new theory is attacked as absurd; then it is admitted to be true, but obvious and insignificant; until finally it is seen to be so important that its former opponents claim that they discovered it. He hoped that *The Meaning of Truth*, published in 1909, would at least help the pragmatist theory of truth from the first into the second stage. Clearly, it did not. My object has been rather different: to ensure that if we reject the theory, we do so for the right reasons; and that, if there is anything of importance to be learned from it, we do not ignore the lesson.*

University of Warwick

REFERENCES

Ayer, A. J. [1968]: *The Origins of Pragmatism*, Macmillan.

Dewey, J. [1901]: 'A Short Catechism Concerning Truth', in *The Influence of Darwin on Philosophy*, Henry Holt.

Dewey, J. [1929]: *Quest for Certainty*. Capricorn Books, 1960.

Dewey, J. [1938]: *Logic, The Theory of Inquiry*, Henry Holt.

Dummett, M. A. E. [1959]: 'Truth', *Proceedings of the Aristotelian Society* 1958–9, and in G. Pitcher (*ed.*) *Truth*. Prentice-Hall, 1964.

Haack, S. [1975]: 'Pragmatism and Ontology: Peirce and James', *forthcoming* in *Revue Internationale de Philosophie.*

Hamlyn, D. [1970]: *Theory of Knowledge*, Macmillian.

Hertz, R. A. [1971]: 'James and Moore: Two Perspectives on Truth', *Journal of the History of Philosophy*, **9**.

Howard, V. [1975]: 'The Pragmatic Maximum' *British Journal for the Philosophy of Science*, **26**, pp. 343–51.

---

* I have been helped by comments made when an earlier version of this paper was read at the York conference on logic and semantics in 1974, and by the students, especially Ken Howse, who followed my course in Recent Anglo Saxon Philosophy at Warwick in 1974–5.

JAMES, W. [1907]: *Pragmatism*, Longman's Green. Page references to Meridian edition, 1955.

JAMES, W. [1909]: *The Meaning of Truth*, Longman's Green. Page references to University of Michigan edition, 1970.

MOORE, G. E. [1908]: 'Professor James' "Pragmatism"'. *Proceedings of Aristotelian Society*, 8. And in *Philosophical Studies*, Routledge, 1922; page references to 1960 paperback edition of *Philosophical Studies*.

MURPHY, M. [1961]: *The Development of Peirce's Philosophy*, Harvard University Press.

O'CONNOR, D. D. [1964]: 'Peirce's debt to F. E. Abbot', *Journal of the History of Ideas*, **25**.

PEIRCE, C. S. [1877]: 'The Fixation of Belief', *Popular Science Monthly*, **12**.

PEIRCE, C. S. [1878]: 'How to Make Our Ideas Clear', *Popular Science Monthly*. **12**.

PEIRCE, C. S. [1908]: Letter to Lady Welby, 1908. All in P. P. Wiener (*ed.*): *Selected Writings*. Dover 1958; page references to Dover edition.

PEIRCE, C. S. [1931–58]: *The Collected Papers of Charles Saunders Peirce*. Harvard University Press.

PERKINS, M. [1952]: 'Notes on the Pragmatic Theory of Truth', *Journal of Philosophy*, **49**.

POPPER, K. R. [1960]: 'Truth, Rationality and the Growth of Scientific Knowledge', in *Conjectures and Refutations*, Routledge & Kegan Paul, 1963.

PRATT, J. B. [1909]: *What is Pragmatism?* Macmillan, 1909.

QUINE, W. V. O. [1951]: 'Two Dogmas of Empiricism', *Philosophical Review*, **60** and in *From a Logical Point of View*, Harper Torchbooks, 1953.

QUINE, W. V. O. [1960]: *Word and Object*, Wiley, 1960.

QUINTON, A. [1966]: 'The Foundations of Knowledge' in B. A. O. Williams and A. Montifiore (*eds.*): *British Analytical Philosophy*. Routledge and Kegan Paul, 1966.

RESCHER, N. [1973]: *The Coherence Theory of Truth*. Oxford University Press.

RUSSELL, B. [1908]: 'James's Conception of Truth', *Albany Review*, **3**, and in *Philosophical Essays*, Allen and Unwin, 1910; references to 1966 revised edition of *Philosophical Essays*.

RUSSELL, B. [1939]: 'Dewey's New Logic', in P. A. Schillp (*ed.*): *The Philosophy of John Dewey*. Tudor, 1939.

SCHEFFLER, I. [1974]: *Four Pragmatists*, Routledge and Kegan Paul.

TARSKI, A. [1931]: 'The Concept of Truth in Formalised Language' in J. Woodger (*ed.*): *Logic, Semantics and Metamathematics*, Oxford University Press, 1956.

Brit. J. Phil. Sci. 27 (1976), 251–274 *Printed in Great Britain*

Discussions

PROGRESS AND DEGENERATION IN THE 'IQ DEBATE': COMMENTS ON URBACH

1 *Hard Core or Fisher.*
2 *The Heritability of Intelligence.*
3 *Critique of Urbach.*
4 *Measuring the Environment.*
5 *Sub-Population Differences.*
6 *Between Groups Heritability in Stature.*

In a recent paper Urbach ([1974]) discusses what he calls progress and degeneration in the IQ debate. Urbach's paper contains in my judgment errors of fact, errors in the interpretation of data, and conceptual misunderstandings. I should like to comment on these, first stating my own position on the general issue. It is this: the controversy about heritability of *IQ*, in particular as it relates to social class and other sub-population differences, is scientifically barren, because the questions are posed in such a way as to admit of no definite, or even 'probabilistic' answer. The 'debate' is being conducted in thoroughly old-fashioned terms which confuse the real issues.

1 HARD CORE OR FISHER

Urbach, following Lakatos ([1970]), believes that heredity-environment issues can be best clarified if one reduces the 'hereditarian' and 'environmentalist' positions to their 'hard core'. It is of course very easy to polarise the nature-nurture controversy in this way, and to do so has a certain elementary didactic value. The point can be illustrated by the familiar example of plant growth. If genetically identical seeds are grown in 'natural' or heterogeneous conditions, all of the variance (differences in growth) will be due to non-genetical or environmental factors. If genetically dissimilar plants are grown in standard conditions, all of the differences will be due to heredity (or more precisely to the differential effects on the growth of those particular plants of that particular set of environmental conditions). Outside the laboratory however the plant breeder or biologist encounters both genetic variability and environmental heterogeneity, and in these circumstances, if he wishes to explain phenotypic variation, he has to attempt to standardise, or bring under control, or vary in specified ways, *some* of the conditions on which his observations are based. *Explaining* his findings means accounting, through multifactor analyses, for as much of the variance as possible, recognising not only that in all empirical work there is bound to be residual, unexplained or so-called error variance, but also that phenotypic growth is always a product of heredity and environment. Even apparently slight changes in either genotype or environment may have profound and unanticipated consequences, and even apparently profound changes may have only slight effects. Scientists since Fisher ([1918]) have viewed the problem of heritability as being one of accounting for variance; and the controversies that have arisen regarding the heritability of biometric attributes concern the proportion of population variance that, under specified conditions, can be ascribed to genetical or environmental factors.

It seems to me that philosophers of science have been slow to consider the implications of the approach to scientific problem solving through methods of multifactor analysis. For example, the contributors to the Lakatos and Musgrave volume *Criticism and the Growth of Knowledge* ([1970]) directed their argument to classical, essentially causal, theories. These have an important, perhaps decisive, place in the explanation of closed, repetitive or stationary systems; but in the open systems studied by social and biological scientists the problems are usually too complex, and the causal factors too numerous and ill-defined, for explanation to be other than statistical. It is characteristic of an open system that the phenomena under investigation are influenced by many factors that can neither be specified nor brought under experimental control. Furthermore, replication is often, in a strict sense, impossible, and can only be approached statistically. In consequence *explanation* is only partial.

There is a further complication: even the fact that we can 'explain' part of the variance doesn't necessarily give us much confidence that the terms we use to do so have any real bearing on the mechanisms which are responsible for the differences. For example the factors which 'account' for mean differences in biometric attributes may do so through their association with other factors which are not included in the analyses: thus there are social class differences in mortality, but social class itself is hardly an explanatory variable; there are regional differences in neural tube defects, but this doesn't mean that these are caused by some teratogenic agent such as mouldy potatoes—on the contrary, as Lowe and Roberts ([1975]) have pointed out, the raised incidence of spina bifida and anencephaly in South Wales may be due to factors which cause women to carry through to term, malformed foetuses which in other environments they would spontaneously abort. More generally, correlation does not imply causation, and any attempt to apportion variance to genetical or environmental factors must be regarded as only a rough first sorting of factors believed to be responsible. In itself such an analysis tells us nothing about mechanisms; moreover any conclusions reached are specific to the population and to the conditions being investigated. It is therefore a matter for empirical study, not just for more sophisticated model building, to see whether similar differences would be found in other circumstances. Not to recognise this is to engage in nineteenth-century mechanist determinism or naive twentieth-century futurology based on the belief that if we could put enough terms into an equation we could explain everything that exists or will exist. In science, as in politics, hard core analyses though often superficially attractive, may be thoroughly misleading.

2 THE HERITABILITY OF INTELLIGENCE

In no field of study has a multifactor approach to the explanation of individual differences been more actively pursued than in psychology; and in no branch of psychology more assiduously than that concerned with intelligence. The reasons are obvious. Mankind is a wild breeding species of extraordinary genetical complexity and mix. Men live in a bewildering variety of environments, the relevant (relevant to *IQ* that is) characteristics of which are largely unknown.

Intelligence is a learned or acquired attribute, a function not a structure, though it has a structural base. And similar environments affect different people in different ways.

No psychologist therefore acts as though there were two hard core research programmes—the one 'hereditarian' and aiming to show that intelligence is *entirely* determined by 'a large number of genes which segregate in accordance with Mendelian principles and which each produce a small, similar and cumulative effect', the other 'environmentalist', asserting that every individual 'inherits identical innate mental capacities'. Instead it would be generally agreed that, as Eysenck ([1973])—whom Urbach quotes but dismisses—has put it, there is 'a single research programme' the aim of which is 'interactionist'. Observed differences in *IQ* are always the result of genetic factors interacting with environmental ones. The aim of 'the research programme' is therefore 'to find out the relative importance of genetic and environmental factors', that is, for a specified population and a specified environment, to partition as much of the *IQ* variance as possible into factors which can reasonably be called genetic and environmental respectively.

An important part of this task is to reduce error, or unexplained, variance to a minimum: to do this could indeed be said to be the crucial problem. Urbach, by ignoring the psychologists' concern with residual variance (and co-variance) I think misunderstands the most significant aspects of the controversy. However, if we could define and measure intelligence unambiguously and with only negligible error, as we can define and measure height, the heritability problem would be greatly simplified. This too is generally recognised: and this is why models of *IQ* heritability have been largely taken over from models applied first to characteristics of physical growth which can be measured unambiguously and reliably

3 CRITIQUE OF URBACH

As Urbach points out, the controversy regarding the heritability of intelligence is in part concerned with the nature and measurement of intelligence, in part with the adequacy and applicability of multifactorial genetical models, and in part with the elucidation of environmental factors which influence development. However, all three sets of issues are much more complex than Urbach would have us believe. I do not propose to discuss the definition and measurement of intelligence: current textbooks on the subject set out the issues clearly. It is however, worth drawing attention to some of the more dubious aspects of Urbach's presentation of the factual and statistical issues. It should be noted that Burt's ([1956]) twin studies on which much of the evidence for very high within-groups heritability rests, must be discounted (Jensen [1974]; Clarke and Clarke [1974]; Kamin [1975]); they are so full of inconsistencies as to make me at least wonder whether the data from the studies which he claims to have carried out, were in fact ever analysed in the way he purports to have done. Heritability estimates of intelligence arrived at by other writers (see Clarke and Clarke [1974]; Rutter and Madge [1975]) have varied considerably, and Burt and Jensen's figure of 80 per cent is by no means universally accepted by psychologists and statisticians who have studied this problem.

Of course for anyone who doesn't adopt a hard core position, this is entirely to be expected (*cf.* the plant example above). But since there is much disagreement over the proportion of the variance which can credibly be attributed to heredity, and since there is no useful taxonomy of relevant environmental characteristics, let alone a metric for measuring them, most of the current controversy becomes a matter of haggling over the question of where to put the ill-defined and unexplained error variance—on the genetic or the environmental side of the equation? In other words, none of the terms in the heredity-environment equation can be unambiguously quantified, and the explanatory power of *all* heritability models is very poor. This makes them largely impervious to criticism and falsification. The general issues are lucidly discussed by Moran ([1974]).

There are a few other errors of fact in Urbach's paper that ought to be pointed out. Normality of distribution is *not* a necessary condition for multifactorially determined biometric attributes. However, as Wechsler ([1944]) noted thirty years ago, the normal curve of *IQ is* a statistical artefact—for example, raw scores on some tests of intelligence, such as the widely used Scottish 11+ intelligence tests, have had a rectilinear distribution which is then normalised for purposes of statistical analysis. It is *not* the case that imbecility is determined mainly by rare, single genes the effect of which is to lower *IQ*. Berg and Kirman ([1959]) and Kirman ([1975]) have shown this fairly clearly: in a series of 645 admissions to hospital of severely mentally handicapped children, data on whom were analysed by Angeli and Kirman, only 32 could be assigned to a 'genetic entity group', the rest being either Down's syndrome (104 cases), environmental (158), unclassified (327), and congenital hydrocephalus (24). I am also highly sceptical of Conway's reported finding that the *IQ* of skilled manual workers in London is declining: a very much better study by Jones ([1959]) does not support this conclusion.

Another largely unconfirmed claim which is accepted uncritically by Urbach concerns regression to the mean over two generations. An exhaustive analysis by McAskie and Clarke ([1975]) of all the published data of *IQ* studies in which *both* parents and children have been tested shows that the classical genetical model of regression in the filial generation to a position half way between the mid-parent *IQ* and the population mean is simply untenable. Similarly Carter and Marshall (to be published) have shown that in the *only* sample on whom data regarding fathers', mothers' and children's heights have so far been analysed, there is *no* regression to the mean from mid-parent height to adult height in the next generation. As McAskie and Clarke ([1975]) point out, most studies in which regression to the mean, from parents to child, has been reported, have measured only one parent, then referred to mid-parent *IQ*, or height or what not. (See Rutter and Madge [1975], for a review of these studies.)

4 MEASURING THE ENVIRONMENT

In the absence of any satisfactory taxonomy of the 'environment' the best we can do, in assessing environmental effects on intelligence, is describe situations in which marked differences in *IQ* are associated with different environmental factors. Current research is concentrated largely on differences in rearing patterns. Thus, to cite some recent studies, Dennis has shown that children brought up in a very depriving orphanage in the Lebanon and adopted at different ages, differ markedly in *IQ* depending upon whether they were adopted before or after the age of two; even more striking were the differences in the later *IQ*s of boys and girls; Barbara Tizard ([1975]) has shown that children in residential institutions in this country differ by 1.5 Standard Deviations in their comprehension of language the differences depending upon the characteristics of the language used by the staff (there were no differences in *IQ* between white and black children in this study); Richardson and others (Richardson [1972] and Tizard [1974]) have shown that clinical malnutrition *and* a poor environment reduce *IQ*s by 20 points, each factor contributing equally and separately to the difference; Yule *et al.* ([1975]) have shown that, among children of Caribbean parents being brought up in London, there is a 9 point difference in non verbal *IQ* between those born in this country and those coming to the country having been born in the West Indies; Gay and Cole ([1967]) and Cole and others ([1971]) have drawn attention to the importance of school attendance in influencing the ability of Kpelle children in Liberia to solve abstract problems; Heber and his associates ([1972]) have shown in an experimental study that there is a 30 point difference in *IQ* between the children of mentally retarded mothers who are brought up in a very stimulating environment as compared with those of similar children brought up in their own homes. We cannot dismiss these, and the less dramatic but nonetheless substantial gains shown in some of the Headstart studies as mere 'anomalies'; and the 'research programme' of which they may be said to form part is perhaps the most exciting work in all contemporary psychology. This is work which is attempting to analyse mechanisms of biological and social development and the specific aspects of 'the environment' which influence it. The genetical counterpart to this work would be the rigorous analysis of multiple gene effects in varying environments. This has not so far been attempted, and it is indeed not yet practicable.

5 SUB-POPULATION DIFFERENCES

In the meantime we are left with the brute factors of individual differences in intelligence (for which no-one denies there is in part a genetical basis) and marked between-groups differences in mean *IQ* (of sub-populations which also differ in numerous largely unspecified ways in the environment in which they live). We simply do not know why these differences exist, and personally I am sceptical of our ability to account for them at the present time. To quote with approval, as Urbach does, Galton's remark that 'so long as we have a plenitude of evidence in favour of the hypothesis of the hereditary descent of talent, we need not be disconcerted when negative evidence is brought against us', as though that put everything right, is to abandon science for dogmatism. As Moran ([1974]) has put it, estimates of heritability 'do not throw any light at all on the difference between different castes (*i.e.* identifiable groups between which there is little or no migration or mating) and attempts to calculate them are quite irrelevant'.

6 BETWEEN GROUPS HERITABILITY IN STATURE

These arguments are however all somewhat general. They can be made more concrete by a further example, using data on 'subpopulation' differences in stature. Recently (Tizard [1975]) I analysed data on height of three groups of London children aged 7 to 12, one group measured in 1909, one group in 1938 and one in 1959. The samples were all very large, and were representative of the London school population at the time. The measurements were carried out in specially conducted surveys undertaken by school doctors. Unlike *IQ* data, the data from these studies are undoubtedly highly reliable, and are by definition valid. They are given in Table 1.

TABLE 1 *Mean heights in centimetres of London boys in 1905, 1938* and *1959*

Age	1909		1938	1959	
	mean	*SD*	mean	mean	*SD*
7.5	114.2	6.04	120.5	123.1	5.69
8.5	119.7	6.12	125.8	129.0	6.07
9.5	124.7	6.23	130.9	134.2	6.51
10.5	129.0	6.56	135.8	139.1	6.48
11.5	133.4	6.69	140.4	144.9	7.20
12.5	137.6	7.33	144.8	149.5	7.65
$\bar{x}$ (ages 7.5–12.5)	126.4	6.51*	133.0	136.6	6.63*

Pooled variance 43.17*; *SD* 6.57*; *computed.

What happens if you apply to these data the type of analysis used by Jensen to support his case that *IQ* differences between blacks and whites are genetic in origin? Two points stand out. First, the phenotypic differences between the 1909 and 1959 samples are much *larger* than the *IQ* differences between blacks and whites, being 10.2/6.57 = 1.55 Standard Deviations apart, compared with one *SD* in *IQ* between blacks and whites. Secondly, since height is a more highly heritable characteristic than *IQ*, the differences *expressed as standard deviations of the genetic component* are perhaps three times as great as those for the *IQ* differences between blacks and whites.[1] The differences in growth (which on average amount to 10.2 cm and are pretty constant from one age group to another) are without question almost entirely due to differences in the environment: better feeding, better material and social conditions and better medical care. Yet if we were simply presented with the London data as height or *IQ* measurements obtained from three 'races', or from two races with an intermediate 'hybrid', the 1938 sample, and if we accepted Jensen's line of reasoning, an environmental explanation of the differences might well seem very strange indeed. What this means is that the 'probabilistic inference' that the higher the within-groups heritability, the more likely it is that between-groups differences have a genetic basis is not supported. The genetical model adds nothing to what we already know, or to what we believe on other grounds; and as Jensen, Burt and others have used height data to exemplify the explanatory power and cogency of the genetic model, the London data may offer a refutation (and I think a decisive refutation) of its theoretical adequacy.

---

[1] The statistical problem, as posed by Jensen ([1973]), is to provide an estimate in standard deviation units, of the amount by which environments would have to differ if all phenotypic between-groups differences were to be attributed to environmental causes. Taking his data, the mean *IQ* of whites = 100, and of blacks = 85, mean difference 15 points. The standard deviation (*SD*) also happens to be 15 points so the variance $(SD)^2 = 225$ points. The within-groups heritability he takes to be 0.80; that is 80 per cent. of the *IQ* variance ($0.80 \times 225 = 180$) is due to heredity and 20 per cent. ($0.20 \times 225 = 45$) to environment. The *SD* of environmental effects is $\sqrt{45} = 6.71$ points. The mean phenotypic difference between blacks and whites is 15 *IQ* points so assuming each group has a within-group heritability of 0.80, it follows that if all of the phenotypic difference (15 points) is due to environmental factors, blacks and whites must differ by 15/6.71 = 2.24 *SD*s in environment as it affects intelligence. Jensen considers this to be implausibly high, and therefore rejects the hypothesis that mean differences between blacks and whites can be accounted for entirely in environmental terms.

Applying the same reasoning to the London height data given in the table and taking Fisher and Burt's figure (cited by Urbach) of 0.95 as the heritability of stature, we have:

D (136.6 − 126.4) = 10.2 cm
Pooled variance = 43.17
Variance of environmental effects (43.17×0.05) = 2.17 cm
SD of environmental effects (square root of 2.17) = 1.47 cm

The mean difference is thus 10.2/1.47 = 6.94 times the 'environmental effects'

It should be added that all the other arguments and data in support of the view that race or sub-population differences in *IQ* are largely genetic in origin can be applied with equal ease to the height data, strengthening the case against the theoretical adequacy of the model. We know, for example, that in spite of the secular increase in height over half a century, social class *differences* remain. Thus whatever their true genetic potential, children in the lower social classes lag behind their environmentally better-off age-mates, though they may equal or surpass children of the higher social classes growing up half a century ago. It is obvious also that the differences cannot be accounted for as 'artefacts' of the method of measurement. Taking other measurements of bodily growth (calf circumference, sub-scapular skin fold, weight, and so on) would provide nothing either to help predict, or explain the reason for, the secular increase in the growth of children which has been going on for more than a century, and of which the London data presented in the table are but a small sample. In short, we cannot draw any firm conclusion as to the extent to which genetic differences are responsible for phenotypic differences between groups growing up in different environments.

Too little attention has been paid to the fact that in the *IQ* debate the term 'heritability' is used in two different senses: within-groups heritability refers to variance, between-groups heritability to differences between means. We cannot argue from the one to the other, and the attempt to do so can only give rise to indeterminate answers which are at best scientifically valueless. The controversy will no doubt continue because of its political implications, but scientifically it is a dead duck.

JACK TIZARD
Institute of Education, University of London

---

expressed in standard deviation units (4.90 *SD* units if as much as 10 per cent of the variance is environmental).

Clearly, if Jensen believes it difficult to account for a standard deviation value of 2.24 for the *IQ* data by an environmental hypothesis, it is very much more difficult to account for a standard deviation value of 6.94, or even 4.90 for the height changes in environmental terms. And even if the heritability coefficient for height in children is less than 0.90, it is nonetheless almost certainly greater than the heritability coefficient of *IQ*, and the observed difference in height between 1909 and 1959 (10.2 cm) expressed in *SD* units is 1.55 times as great (10.2/6.57 = 1.55) as the corresponding difference of one *SD* in *IQ*, between blacks and whites.

REFERENCES

BERG, J. and KIRMAN, B. [1959]: 'Some Aetiological Problems in Mental Deficiency', *British Medical Journal*, 2, pp. 848–52.

BURT, C. and HOWARD, M. [1956]: 'The Multifactorial Theory of Inheritance and Its Application to Intelligence', *British Journal of Statistical Psychology*, 9, 95–131.

CARTER, C. O. [1975]: Paper given to the Society for the Study of Human Biology, April, 1975.

CLARKE, A. M. and CLARKE, A. D. B. [1974]: *Mental Deficiency: The Changing Outlook*. London: Methuen.

COLE, M., GAY, J., GLOCK, J. and SHARP, D. W. [1971]: *The Cultural Context of Learning and Thinking*. New York: Basic Books.

CONWAY, J. [1959]: 'Class Differences in General Intelligence II', *British Journal of Statistical Psychology*, 12, pp. 5–14.

DENNIS, W. [1973]: *Children of the Crèche*. New York: Appleton-Century-Crofts.

EYSENCK, H. J. [1973]: *The Inequality of Man*. London: Maurice Temple Smith.

FISHER, R. A. [1918]: 'The Correlation between Relatives on the Supposition of Mendelian Inheritance'. *Transactions of the Royal Society of Edinburgh*, 52, pp. 399–433.

GAY, J. and COLE, M. [1967]: *The New Mathematics and an old Culture: Study of Learning among the Kpelle of Liberia*. New York: Holt Rhinehart and Winston.

HEBER, R., GARBER, H., HARRINGTON, S., HOFFMAN, C. and FALENDER, X. [1972]: *Rehabilitation of Families at Risk for Mental Retardation*. Madison, Wisconsin: December 1972. Progress Report, Rehabilitation Research and Training Center in Mental Retardation, University of Wisconsin.

JENSEN, A. R. [1973]: *Educability and Group Differences*. London: Methuen.

JENSEN, A. R. [1974]: 'Kinship Correlations Reported by Sir Cyril Burt', *Behaviour Genetics*, 4, pp.1–28.

JONES, S. [1959]: 'A Statistical Study of the Wechsler Intelligence Scale for Children', University of London unpublished Ph.D. thesis.

KAMIN, L. J. [1974]: *The Science and Politics of I.Q.* Potomac, Maryland: Lawrence Erlbaum Associates.

KIRMAN, B. [1975]. 'Some Causal Factors', in B. Kirman, and J. Bicknell, *Mental Handicap*, pp. 67–90. Edinburgh: Churchill.

LAKATOS, I. [1970]: 'Falsification and the Methodology of Scientific Research Programmes', in I. Lakatos and A. Musgrave (*eds.*): *Criticism and the Growth of Knowledge*, pp. 91–196. Cambridge: Cambridge University Press.

MCASKIE, M. and CLARKE, A. M. [1976]: 'Natural Parent–Offspring Resemblances in Intelligence: Theories and Evidence', *British Journal of Psychology, forthcoming*.

MORAN, P. A. P. [1974]: 'The Estimation of Psychological Genetic Differences Between Classes and Races', *British Journal of Psychiatry*, 124, pp. 228–92.

RICHARDSON, S. A. [1972]: *Nutrition, the Nervous System and Behaviour*, Pan American Health Organization (World Health Organization), Scientific Publication No. 251.

ROBERTS, C. J. and LOWE, C. R. [1975]. 'Where have all the Conceptions Gone?', *The Lancet*, 1 March, pp. 498–9.

RUTTER, M. and MADGE, N. [1975]: *Cycles of Disadvantage: A Review of Research*, Social Science Research Council Publication.

TIZARD, B. [1975]: 'Varieties of Residential Nursery Experience', in R. V. Clarke, I. Sinclair, J. Tizard (*eds.*): *Varieties of Residential Experience*. London: Routledge & Kegan Paul.

TIZARD, J. [1974]: 'Early Malnutrition, Growth and Mental Development in Man', *British Medical Bulletin*, 30, No. 2.

TIZARD, J. [1975]: 'Race and IQ: the Limits of Probability', *New Behaviour*, 24 April, pp. 6–9.

URBACH, P. [1974]: 'Progress and Degeneration in the "IQ Debate"', *British Journal for the Philosophy of Science*, 25, pp. 99–135, 235–59.

WECHSLER, D. [1944]: *The Measurement of Adult Intelligence*, Baltimore: Williams and Wilkins.

YULE, W., BERGER, M., RUTTER, M. and YULE, B. [1975]: 'Children of West Indian Immigrants: Intellectual Performance and Reading Attainment', *Journal of Child Psychology and Psychiatry*, pp. 1–17.

ON EINSTEIN'S SECOND POSTULATE

In a recent note Kenneth Schaffner has argued[1] that Einstein shifts from one form of the second postulate (or 'light postulate') to another in the course of his classic 1905 paper. The second postulate formulated by Einstein in his introductory section (hereafter 'the second postulate$_1$'),[2] Schaffner takes to be the '*I.L.V.* hypothesis', according to which the velocity of light is independent of motion of the light source.[3] That which Einstein gives as his second postulate in his second section (hereafter 'the second postulate$_2$'), Schaffner claims is the stronger '*C.L.V.* hypothesis', which asserts that the speed of light is a constant which is independent not only of motion of the light source, but also of the state of motion of the observer's frame of reference. In an earlier article to which Schaffner refers, Adolf Grünbaum has similarly interpreted the second postulate$_2$ to be the *C.L.V.* hypothesis.[4]

I believe this exegesis of the second postulate$_2$ to be mistaken for the following reasons. In section 3 of his paper, Einstein's derivation of his space and time transformations utilises the following procedure:

> . . . we can easily determine the quantities ξ, η, ζ by expressing in equations that light (as required by the principle of the constancy of the velocity of light *in combination with* the principle of relativity) is also propagated with velocity c when measured in the moving system.[5] (Italics mine.)

Since the *C.L.V.* hypothesis *alone* entails that light is propagated with velocity c in *all* inertial systems, it is unlikely that Einstein intends 'the principle of the constancy of the velocity of light' to be the *C.L.V.* hypothesis, since this would make his invocation of the principle of relativity superfluous. But it is only the second postulate$_2$, not the second postulate$_1$, that Einstein explicitly designates 'the principle of the constancy of the velocity of light'.[6] Hence the above passage is strong evidence that Einstein intends the second postulate$_2$ to be the *I.L.V.* hypothesis, not the *C.L.V.* hypothesis.

Moreover, on the next page Einstein writes:

> We have now to prove that any ray of light, measured in the moving system, is propagated with the velocity c, if, as we have assumed, this is the case in the stationary system; for we have not yet furnished the proof that the principle of the constancy of the velocity of light is compatible with the principle of relativity.[7]

Since, as Schaffner points out, there is not even a *prima facie* incompatibility between the *C.L.V.* hypothesis and the principle of relativity, this passage reinforces our exegetical conclusion: contra Schaffner and Grünbaum, Einstein's second postulate$_2$ is not the *C.L.V.* hypothesis, but rather the *I.L.V.* hypothesis, whose compatibility with the principle of relativity must accordingly be proven.

---

[1] Schaffner [1974]: my comments focus on pp. 54–5.

[2] Einstein [1905], p. 38.

[3] When treating the *I.L.V.* hypothesis as a postulate of Einstein's theory, Schaffner is undoubtedly stripping it of the aether-theoretic connotation (that light has the velocity c 'only in the aether rest frame') which, on p. 54, he misleadingly characterises as *part of* the *I.L.V.* hypothesis.

[4] Grünbaum [1964], p. 1407.

[5] Einstein [1905], p. 45.

[6] Einstein [1905], p. 41.

[7] Einstein [1905], p. 46.

Also, there is a puzzle to be raised concerning Einstein's second postulate$_1$. As Schaffner italicises, Einstein remarks that this postulate is 'apparently irreconcilable' with the principle of relativity. It is therefore tempting to agree that it must be the *I.L.V.* hypothesis. But in the wording of the postulate Einstein says: '... light is *always* propagated in empty space with a definite velocity *c* ...'.[1] The word 'always', which does not occur in the second postulate$_2$, seems in this context to bear no interpretation except 'in all inertial frames', which would make this postulate the *C.L.V.* hypothesis. Pulled in one direction by 'apparently irreconcilable' and in another by 'always', perhaps we can safely conclude only that Einstein was a bit careless in articulating this postulate.

Because of its exegetical importance, it is necessary to be clear on why Schaffner is correct in asserting that it is the *I.L.V.* hypothesis, but not the *C.L.V.* hypothesis, that is 'apparently incompatible' with the principle of relativity, even though, as Schaffner also points out, the *C.L.V.* hypothesis entails the *I.L.V.* hypothesis. The ostensible logical oddness of this evaporates when one sees the following. The *I.L.V.* hypothesis is apparently incompatible with the principle of relativity because it appears to entail that the velocity of light depends upon the motion of the frame of reference from which it is measured (hereafter '*D.L.V.*'), which contradicts the principle of relativity.[2] The *C.L.V.* hypothesis, however, goes beyond the *I.L.V.* hypothesis precisely by *denying* this apparently entailed consequence of the *I.L.V.*, for it asserts both *I.L.V.* *and not-D.L.V.* There is thus no apparent incompatibility between the *C.L.V.* hypothesis and the principle of relativity: the question instead is whether the *C.L.V.* hypothesis is *internally consistent.* If it is not, it is incompatible with the principle of relativity only in the sense—surely trivial in this context—in which a logical contradiction can be said to be incompatible with everything and anything.

Einstein's thesis is then that the *I.L.V.* hypothesis *appears* to entail *D.L.V.* because we tacitly assume the classical Galileian transformations, and that there is an alternative set of acceptable transformations by which the *I.L.V.* hypothesis does not entail *D.L.V.* and so is consistent with the principle of relativity, and by which the *C.L.V.* hypothesis is internally consistent.[3] Einstein's originality, as Grünbaum has argued, lies in his undergirding of these 'Lorentz transformations' with a cogent reanalysis of simultaneity and a relational view of space and time.[4]

I submit the above exegetical contentions on the assumption that the issues they raise concerning Einstein's epoch-making paper are, in their own right, worth clearly resolving. I leave as an open question whether, in addition, they throw further light on the larger philosophical issues with which Grünbaum and Schaffner are concerned.[5]

STEVE WYKSTRA
University of Pittsburgh

---

[1] Einstein [1905], p. 38.

[2] The contradiction arises, Adolf Grünbaum pointed out to me, on Einstein's assumption that the velocity of light is law-like, not just fact-like.

[3] See Einstein [1917], pp. 9–34 for what is perhaps Einstein's most lucid presentation of this.

[4] Grünbuam [1973], pp. 400–409.

[5] I wish to thank Kenneth Schaffner and Adolf Grünbaum for their comments on an earlier version of this paper.

REFERENCES

EINSTEIN, A. [1905]: 'Zur Electrodynamic bewegter Körper', *Annalen der Physik*, **17**, pp. 891–921. Page references are to English translation in A. Einstein *et al.*, *The Principle of Relativity*, pp. 35–65.

EINSTEIN, A. [1917]: *Über die Spezielle und die Allgemeine Relativitätstheorie, Gemeinverständlich*. Page references are to the English translation *Relativity: the Special and General Theory*.

GRÜNBAUM, A. [1964]: 'The Bearing of Philosophy on the History of Science, *Science*, **143**, pp. 1406–12.

GRÜNBAUM, A. [1973]: *Philosophical Problems of Space and Time*. Second revised edition.

SCHAFFNER, K. [1974]: 'Einstein versus Lorentz: Research Programmes and the Logic of Comparative Theory Evaluation', *The British Journal for the Philosophy of Science*, **25**, pp. 45–78.

SOME NEW ASPECTS OF RELATIVITY: REMARKS ON ZAHAR'S PAPER*

In an excellent paper published in this *Journal*, Zahar[1] has shown that

(1) In the earliest days of relativity theory Einstein's theory had not superseded Lorentz's.

(2) Supersession took place only in 1915.

The greater part of his paper gives a detailed argument in favour of (1), whereas (2) is dealt with in only a few paragraphs.

Whereas Feyerabend[2] briefly criticised (2), both Schaffner[3] and Miller[4] concentrated their polemical analyses against (1). However, the criticisms made by both Schaffner and Miller are wrong as regards their central points. Indeed it seems to me that their mistakes are the result of a misunderstanding, both of the source material, and of Zahar's paper. I shall deal with each of these points in turn.

1 Both Schaffner and Miller state that the Lorentz–Fitzgerald Contraction Hypothesis (*L.F.C.*) is *ad hoc*, whereas—and this is particularly emphasised by Schaffner—Einstein's light postulate possessed considerable independent support. For instance the constancy of the velocity of light (*C.L.V.*) is confirmed by the Michelson–Morley experiment. But the *L.F.C.* was not *ad hoc*.

The *L.F.C.* was formulated to explain the result of the Michelson–Morley experiment. It could be considered *ad hoc* only if it had no independently testable consequences, or, at least, if there were no independent tests available at the time. However, such tests were available.

One of them is the Trouton–Noble experiment (carried out in 1903). As Butler[5] has shown, the null result of the Trouton–Noble experiment is an immediate consequence of the *L.F.C.* alone. Other explanations given at the time failed. For example, Laue[1] attempted an explanation in 1911 on the basis of the energetical balance but a satisfactory explanation in these terms was given by Butler only in 1968.[2]

---

* I would like to thank Gregory Currie for his help in improving the English throughout this paper.

[1] Zahar [1973].

[2] Feyerabend [1974].

[3] Schaffner [1974].

[4] Miller [1974].

[5] Butler [1968], pp. 940–1.

It was widely believed that, since the prediction of a positive result is due to the non-zero electromagnetic force-couple which is obtained using the Galilean transformations, the null result would be explained by relativistic physics. But as Strnad has shown,[3] the prediction for the force-couple does not become zero when we use relativity physics. Indeed the predicted value for the force-couple is exactly the same as that obtained using classical physics. Moreover, as I have shown elsewhere,[4] the null result is independent both of the transformation equation for the charge density, and of the invariance or otherwise of the electric charge.

On the other hand, the *L.F.C.* gives a simple explanation of the Trouton–Noble experiment. Briefly it is this. A sphere undergoing translation becomes ellipsoid, and the electromagnetic force couple in superposition with the electrostatic forces acts just in the direction perpendicular to the ellipsoid's surface, being thus compensated by the rigidity of the material of the ellipsoid (or of the connecting rod used in the experiment). The null result of the Trouton–Noble experiment is thus an independent test for the *L.F.C.*

Further independent support for the *L.F.C.* known at that time is the law of the variation of mass with velocity (experimentally verified in 1914[5]). The mass formula can be derived by using the law of the conservation of mass and the law of the conservation of momentum, together with the relativistic addition theorem for velocities. But the latter can be derived, not only by using relativity theory, but also by using the invariance of Maxwell's equations with respect to an aether-based transformation, containing transversal and longitudinal contractions of lengths.[6] Because of this derivation, the mass formula can be considered as further independent support for the theory that lengths are physically contracted.

I should like to mention in passing that both Miller and Schaffner use the fact that the Thorndike–Kennedy experiment was not carried out until 1932, as further evidence against Zahar's views. But it should be noted that this experiment is logically equivalent to the Michelson–Morley experiment, and therefore cannot possibly falsify the *L.F.C.*, contrary to what has been repeatedly stated by several authors. I hope to deal with this question elsewhere.

2 Schaffner, discussing experiments of the Thorndike–Kennedy type, argues that 'There is a time measurement while the crucial movement is taking place.'[7] This suggestion, though at first glance attractive, is untrue. Neither in the Michelson–Morley nor the Thorndike–Kennedy experiment does the measurement of time play any role. That experiments of this type cannot give any evidence of time dilatation has been shown by Ives[8] and Palacios.[9] Given Maxwell's equations, experiments of the Michelson–Morley and Thorndike–Kennedy type falsify only the old Newtonian theory, but cannot give any

---

[1] Laue [1911].

[2] Butler, *op. cit.*

[3] Strnad, [1970], p. 64.

[4] Podlaha [1974*c*].

[5] Neumann [1914].

[6] Podlaha [1974*a*], [1974*c*] and [1975*a*].

[7] Schaffner [1974], p. 66.

[8] Ives and Stilwell [1937], pp. 179–80.

[9] Palacios [1961], pp. 201–7.

decision between Einstein and Lorentz or other theories like those of Palacios or Gordon where no time dilatation appears.[1] (In spite of this fact, 'crucial' experiments of this type which are supposed to decide between Einstein and Lorentz still appear.[2])

3 Miller's statement that '. . . a theory of light containing Lorentz's ether cannot account exactly for the optical Doppler effect', is wrong. Miller could have easily avoided this erroneous judgment if he had carefully studied the series of papers by Ives, appearing in the years 1937–48 in the *Journal of the Optical Society of America* or if he studied at least the paper by Robertson of 1949.[3]

4 Also Miller states that 'The inclusion by Lorentz of true length and time quantities which remain unknown to the observer in an inertial system, renders unclear the meaning of the contraction of a rod at rest relative to S by an observer in S' . . .'.[4] This is wrong. On the contrary, as I have shown in detail elsewhere,[5] the *real* contraction of a rod as well as the *real* slowing down of the clock have a quite clear physical meaning, and make it possible to build a quite consistent theory of space and time.

In my [1975*a*] and [1974*b*] I have also shown that all relativistic phenomena have their origin only in the contraction of lengths and in the slowing down of clocks which are as *real* as the contraction of a material rod due to the decrease of temperature.

5 Further, I feel that the Einstein *C.L.V.* is not merely *ad hoc* but that it is an empirically empty statement. It cannot be compared with, for instance, the laws of thermodynamics. In this I sharply disagree with Schaffner. Let me analyse this point in detail.

What do we really know about the spreading of light?

(1) As it is well known, Fresnel already in 1828, formulated the principle that the velocity of spreading of a wave is independent of the velocity of the source. This principle is obvious, as can easily be seen by, for instance, throwing stones at the surface of water. Because light also has wave properties, the same principle is valid for light. Einstein's theory was not necessary for the recognition of this, Fresnel had pointed it out one century earlier. Actually, the experimental corroboration of this postulate is not easy, but in principle it can be made.[6] That the velocity of the wave is independent of that of the source says nothing about the quantitative value of the velocity of light. What is the value of the velocity of light in our laboratory? Or in our solar system? In every textbook of physics we read that the velocity of light is 300,000 km/sec and at least two methods of carrying out this measurement are given: those of Fizeau and Roemer. It is still generally believed that Roemer's method measures the value of the one way velocity of light. What Roemer and Fizeau actually measured was the round-trip velocity of light.

[1] Palacios [1957] and Gordon [1962]. See also Podlaha [1971].

[2] Shamir and Fox [1969].

[3] Robertson [1949].

[4] Miller [1974], p. 43.

[5] Podlaha [1975*a*] and [1975*b*].

[6] Alväger *et al.* [1964].

(2) The fact that the velocity of light in one direction cannot be measured was already well known to Ives, but his papers have been either ignored or wrongly interpreted. Recently Karlov[1] has shown in detail, the impossibility of the measurement of the one way velocity of light by Roemer's method.

It is simply the case that the one way velocity of light cannot be measured at all, and this statement should be explicitly formulated as a law of nature before taking any further steps.[2] It is this principle of the impossibility of the measurement of the one way velocity of light—and not the Einsteinian requirement that the velocity of light is the same in all directions—which is analogous to principles like the second law of thermodynamics.

A direct *consequence* of this principle is that no absolute velocity can be measured and relative velocities can be *measured* only with a lack of sharpness in *principle*. The uncertainty of our measurement depends linearly on the value of the measured relative velocity.[3]

These considerations may appear tedious and elementary, but let me emphasise that to my knowledge, Einstein never made an unconditional and explicit statement about the impossibility of the measurement of the one way velocity of light. In respectable journals, arguments and *experiments* dealing with the one way velocity of light still appear.[4]

If Einstein did formulate the principle of the impossibility of the measurement of the one way velocity of light, he could have avoided a lot of unnecessary reasoning about the impossibility of the measurement of absolute motion. Indeed, the impossibility of the measurement of absolute motion is a direct consequence of the impossibility of the measurement of the one way velocity of light.

(3) *Having formulated this principle* we can either

(*a*) make the convention to consider by definition and for the sake of simplicity only, the values of the velocity of light in the frame S to be equal to c in all directions, where c is the round trip light velocity, or

(*b*) we can choose quite arbitrarily a rule about how to synchronise distant clocks in the frame S. One possible choice being the Einstein synchronisation according to which the time on the clock B is given by the formula:

$$t_B = \frac{t_A + t_A^*}{2} \tag{1}$$

where t_A is the time of the light emission from A, t_B the time of the light arrival at B and t_A^* the time of the light arrival back at A.

Now according to the definition of velocity

$$C_{AB} = \frac{X_B - X_A}{t_B - t_A} \tag{2}$$

$$C_{BA} = \frac{X_A - X_B}{t_A^* - t_B}\,. \tag{3}$$

If we chose (*a*) then, as a consequence of the chosen convention, we *have to* synchronise the clocks according to (1) so as not to be in disagreement with (2) and (3). The choice (*a*) being once made, there is no freedom as regards the choice of the type of synchronisation.

---

[1] Karlov [1970].

[2] Althaus, Podlaha and Bensch [1975].

[3] *Ibid.*

[4] See for example, Trimmel *et al.* [1973], or Cialda [1972].

If (*b*) is chosen, and we synchronise the clock according to the Einstein rule (1) then, as a consequence of the chosen synchronisation we obtain

$$C_{AB} = -C_{BA}. \tag{4}$$

The Einstein synchronisation (1) and the equality of the one way light velocities in opposite directions (4) are two equivalent *conventions*, I repeat, *both being nothing but conventions.*

Let us further note, that, *with respect to those considerations, we are still in the frame S and we make no statements about the velocities in any other frame, say S′.*

Thus it can be seen that the equality of the light velocities in different directions is in no way testable, and has no genuine empirical content.

No matter how tedious these considerations about so elementary and simple facts may seem, every detail is here very important. Einstein failed to consider just these details. Let us see how this happened.

In 1905, Einstein postulated:

> . . . light is always propagated in empty spaces with a definite velocity c which is independent on the state of motion of the emitting body.
>
> $$c = \frac{\text{light path}}{\text{time interval}}$$
>
> Time interval is to be taken in the sense of the definition of simultaneity.[1]

Einstein did not separate the statement about the independence of the light velocity of that of the source from the statement that light is propagated in opposite directions in the frame S with the same velocity.

Further, he mentioned, that the character of his synchronisation, and thus of simultaneity, is *definitional* but he omitted to point out that the constancy of the velocity in various directions is merely a consequence of this definition. Instead he formulated the constancy of the velocity of light as a postulate ascribing the value c to the one way velocity of light. This gives the impression that it is not a pure convention, but an axiom which could be tested. But this is not the case.

Schaffner further writes: 'The definitional component in [Einstein's] reconstruction of the concept was required because the velocity of light in one way direction cannot be ascertained in any empirical manner prior to a definition of distant simultaneity.'[2]

Formulations like this which are not actually false, but where the *whole* truth is concealed are the most misleading ones. This statement of Schaffner's would have us believe that everything would be just fine if we first decided how to synchronise distant clocks. But, in the special theory of relativity it is impossible to determine distant simultaneity, either by reference to a natural law or one of its consequences. The synchronisation of distant clocks can only be a matter of choice. The value of the velocity of light in one direction, being dependent only on the prior synchronisation of the clocks, is therefore no natural constant, no 'New Absolute': whether it is defined in a straightforward way, or by the choice of a particular synchronisation procedure is irrelevant. It is not a natural law in the special theory of relativity.

---

[1] Einstein [1905], p. 895.

[2] Schaffner [1974], p. 57.

Lorentz was not faced with this difficulty. For him, in a frame at rest with respect to the aether, the propagation of light with velocity c in all directions is not a convention but a real natural phenomenon which has its explanations in the existence of the aether. Therefore, in a theory of the aether and only in a theory of the aether, the constancy of the velocity of light, at least in one frame S, can be given a realistic interpretation. As regards this point, therefore, Lorentz's theory is still superior to that of Einstein.

6 Only now as a further consideration can the quite independent question be dealt with as to what happens to the velocity of light (already defined in some way in S) if the situation is observed from another inertial frame S'. This question I intend to deal with in another paper. My intention here was only to point out some errors involved in the arguments against Zahar's paper.

One final remark: I feel that the statement by Miller,[1] that not only Poincaré but also Einstein in 1905 recognised the group character of the Lorentz transformation, is wrong. I cannot find any statements or discussions of these group properties in Einstein's 1905 paper. On the contrary, I find there are some errors in connection with this subject. But this point has been discussed in detail elsewhere.[2]

M. F. PODLAHA
Prague and *Georg-August-Universität, Göttingen*

REFERENCES

ALVÄGER, T., FARLEY, F. J. M., KJELLMAN, J., and WALLIN, I. [1964]: 'Test of the Second Postulate of Special Relativity in the GeV Region', *Physics Letters* **12**, pp. 260–62.

BECHERER, A. H. [1909]: 'Die Experimentelle Bestätigung des Relativistätsprinzips', *Annalen der Physik*, **28**, pp. 513–36.

BUTLER, J. W. [1968]: 'On the Trouton–Noble Experiment', *American Journal of Physics*, **36**, pp. 936–41.

CIALDA, R. [1972]: 'A New Test of the Second Postulate of Special Relativity Sensitive to First Order Effects', *Lettera al Nuovo Cimento*, **4**, pp. 821–5.

EINSTEIN, A. [1905]: 'Zur Electrodynamik bewegten Körper', *Annalen der Physik*, **17**, pp. 891–921.

FEYERABEND, K. [1974]: 'Zahar on Einstein', *The British Journal for the Philosophy of Science*, **25**, pp. 25–8.

GORDON, C. N. [1962]: 'An Alternative Deduction from the Michelson–Morley Experiment', *Proceedings of the Physical Society*, **80**, pp. 569–92.

IVES, H. E. and STILWELL, G. R. [1937]: 'Graphical Exposition of the Michelson–Morley Experiment', *Journal of the Optical Society of America*, **27**, pp. 177–80.

KARLOV, V. [1970]: 'Does Römer's Method Yield a Unidirectional Speed of Light?', *Australian Journal of Physics*, **23**, pp. 243–53.

LAUE, M. [1911]: 'Bemerkungen zum Hebelgesetz in der Relativistätstheorie', *Physikalische Zeitschrift*, **12**, pp. 1008–10.

MILLER, A. [1974]: 'On Lorentz's Methodology', *The British Journal for the Philosophy of Science*, **25**, pp. 29–45.

NEUMANN, G. [1914]: 'Die Träge Masse Schnell Bewegter Elektronen', *Annalen der Physik*, **45**, pp. 529–79.

---

[1] Miller [1974], p. 44.

[2] Podlaha [1974*b*] and [1975*b*].

PALACIOS, J. [1957]: 'Revision de la Teoria de la Relatividad', *Revista de la Real Academia de Ciencias Exactas Fisicas y Natureles de Madrid*, **51**, pp. 56–73, 165–73.

PALACIOS, J. [1961]: 'A Reappraisal of the Principle of Relativity as Applied to Moving Interferometers', *Revista de la Real Academia de Ciencias Exactas Fisicas y Naturales de Madrid*, **55**, pp. 201–7.

PODLAHA, M. [1969]: 'Lorentz Theory, Palacios Theory and Interferometrical Experiments', *Il Nuovo Cimento*, **64B**, pp. 181–7.

PODLAHA, M. [1971]: 'Reply to Shamir's Paper: New Experimental Test on Special Relativity', *Lettere al Nuovo Cimento*, **2**, pp. 601–2.

PODLAHA, M. [1974*a*]: 'Uberlegungen zum Verhältnis von Newtonischen Physik zur Spezielten Relativitätstheorie', *Report D.F.G.*, University of Göttingen.

PODLAHA, M. [1974*b*]: 'The Axiomatic Foundations of the Theory of Special Relativity—Reply to Stiegler', *International Journal of Theoretical Physics*, **11**, pp. 69–72.

PODLAHA, M. [1974*c*]: 'Analysis of Invariance of Maxwell's Equations with Respect to Assymmetrical Linear Transformations of Voigt–Palacios–Gordon Type', *Indian Journal of Theoretical Physics*, **22**, pp. 73–9.

PODLAHA, M. [1975*a*]: 'Length Contraction and Time Dilatation in the Special Theory of Relativity—Real or Apparent Phenomena?', *Indian Journal of Theoretical Physics*, **23**, pp. 69–75.

PODLAHA, M. [1975*b*]: 'An Explanation of the Lorentz Transformation by Non-Relativistic Premises', *Report D.F.G.*, University of Göttingen.

PODLAHA, M., ALTHAUS, B. and BENSCH, H. [1975]: 'Zur Problematik der Geschwindigkeitsmessung', *Philosophia Naturalis*, in print.

ROBERTSON, H. P. [1949]: 'Postulate versus Observation in the Special Theory of Relativity', *Review of Modern Physics*, **21**, pp. 378–82.

SCHAFFNER, K. F. [1974]: 'Einstein Versus Lorentz: Research Programmes and the Logic of Comparative Theory Evaluation', *The British Journal for the Philosophy of Science*, **25**. pp. 45–78.

SHAMIR, J. and FOX, R. [1969]: 'A New Experimental Test of Special Relativity', *Il Nuovo Cimento*, **62B**, pp. 258–64.

STRNAD, J. [1970]: 'A Note on the Trouton–Noble Experiment', *Contemporary Physics*, **11**, pp. 59–64.

TRIMMEL, W. S. N., BAIERLEIN, R. F., FALLER, J. E. and HILL, H. A. [1973]: 'Experimental Search for Anisotropy in the Speed of Light', *Physical Review D2*, **8**, p. 3321.

ZAHAR, E. [1973]: 'Why did Einstein's Programme Supersede Lorentz's?', *The British Journal for the Philosophy of Science*, **24**, pp. 95–123, 223–62.

UNPREDICTABILITY: A REPLY TO CARGILE AND TO BENDITT AND ROSS

1 My 'Unpredictability of Free Choices'[1] has given rise to much more than the normal amount of opposition. When so many philosophers feel that something is faulty with a given thesis, then whether or not they can articulate what the fault is and even if the arguments they marshal against it are rather shaky, it could be a sign that they are correct in sensing a deep defect in the reasoning. On the other hand, it could also simply be my conclusion which so upsets people, since they fear it may lead to a picture of human nature repugnant to them. Although I am loth to upset people, I believe that the latter rather than the former is the case, and so I shall argue in defense of my position.

2 In this note I shall concentrate mainly on answering the attempted attacks of Cargile,[2] and Benditt and Ross.[3] They all agree on one thing: my argument

---

[1] Schlesinger [1974].

[2] Cargile [1975].

[3] Benditt and Ross [1975].

that it is best to take both boxes because this is in accordance with what would be the advice of a sufficiently intelligent and well-informed well-wisher, is wrong. Now before attempting to demonstrate that they have produced no valid arguments to support their position, I should like to show that even if my 'well-wisher argument' was completely fallacious there still would be no grounds for their contention that the agent should fare best if he chooses to take box II only.

What are the alleged reasons for claiming that in order to maximise his gain the agent should take box II only? Supposedly we are given an infallible Predictor who if he predicts at t_0—twenty-four hours before t_1—that at t_1 the agent takes box II only, he places $\$M$ in that box, otherwise he leaves it empty. Now how can we in practice be *given* that someone is an infallible predictor in the future as well as in the past? The answer must be that what we could be given is very strong inductive evidence that this is so. What would be the nature of such evidence? It could be suggested for instance that the Predictor in Newcomb's story has played the game one million times before and in the 500,000 cases in which players with widely different backgrounds and temperaments using a great variety of arguments have chosen to take box II only as well as in the 500,000 cases in which they ended up taking both boxes, has without a single exception correctly anticipated their choices. Do we then indeed have strong inductive evidence that we are faced with a Predictor with whose powers the agent must reckon and thus restrain himself and refrain from taking box I so as to ensure, or to say the least increase the chances, that he finds $\$M$ in box II? Not at all! Let me explain.

It is well known that in empirical reasoning it is not the case that it depends exclusively on the nature of a given piece of evidence and on the nature of a given hypothesis whether the former supports the latter. But whether a certain observation confirms a given hypothesis depends also on the theories we hold and in the context of which the confirmation is supposed to take place. I shall presently describe a theory which is by no means absurd and in the context of which the spectacular success of the Predictor provides no indication whatever that it may be better for the agent to choose box II only.

Let: C_1 = The agent chooses to take box II only.
C_2 = The agent chooses to take both boxes.
P_1 = The Predictor predicts that C_1.
P_2 = The Predictor predicts that C_2.
T_1 = The agent has a tendency to do C_1.
T_2 = The agent has a tendency to do C_2.
D_1 = The Predictor diagnoses that T_1.
D_2 = The Predictor diagnoses that T_2.

Now it is maintained that at least 24 hours before anyone makes a choice he has a distinct, and in principle discernible, tendency to end up making that choice. Within that period of time the agent cannot change the tendency he has ingrained in him but by an act of free will he can behave contrary to his tendency. It is also asserted that:

$$T_1 \leftrightarrow D_1 \quad \text{and} \quad T_2 \leftrightarrow D_2$$

which means that our co-called 'Predictor' is really a perfect diagnostician of tendencies and at t_0 can without fail recognise the tendency the agent has to choose at t_1.

Also:

$$D_1 \leftrightarrow P_1 \quad \text{and} \quad D_2 \leftrightarrow P_2$$

i.e. the 'Predictor' always bases his prediction of what the agent will actually choose on his correct diagnosis of what tendency he has to choose. From this it follows of course that:

$$T_1 \leftrightarrow P_1\ (\alpha) \quad \text{and} \quad T_2 \leftrightarrow P_2\ (\alpha')$$

There is no reason why we should not also maintain that:

$$p(C_1/T_1) = p(C_2/T_2) = 0.99 \ldots 99$$

i.e. that the probability of the agent doing C_i when it is a fact that T_i is extremely high; the vast majority of people who have the tendency to make a given choice end up making that choice. Also, of course:

$$p(C_1/T_2) = p(C_2/T_1) = 0.00 \ldots 01$$

because very few agents act contrary to their tendency.

In order to represent matters graphically, let me assert that any player who plays Newcomb's game is in one of the following four classes:

	T_1	T_2
C_1	a: very large $p(P_1) = 1$	c: very small $p(P_1) = 0$
C_2	b: very small $p(P_1) = 1$	d: very large $p(P_1) = 0$

Since class a is much larger than class c, then, if all we are given is that our agent is in a class characterised by C_1 then we know he is much more likely to be in a than in c or characterised by T_1 rather than by T_2 or (because of (α) and (α') by P_1 rather than P_2). Also when we are given no more than that the agent is in class C_2 then he is much more likely to be in d than in b. It follows therefore that:

$$p(T_1/C_1) > p(P_2/C_1) \quad \text{and} \quad p(P_2/C_2) > p(P_1/C_2)$$

Thus the probability that the Predictor will make the right prediction is vastly greater than that he will err in his prediction. This fits very well with our past observation of the Predictor's performance.

We observe however, that vertical movement only is possible for the agent, that is, he can move from a to b or b to a and from c to d or d *to* c but not, for instance from c to a or d to a. Thus if it is the case that the agent is in the class characterised by T_2 then by going against his tendency and doing C_1, all he can achieve to transfer himself from d to c—and not from d to a—which is of course of no use to him at all since the Predictor who always bases his prediction on his diagnosis of the tendency the agent has, will still predict that C_2 and leave box II empty. Thus he will give up the \$1,000 of box I without in the slightest increasing his chances for receiving \$$M$. Also, if it is a fact that the agent is in a class by T_1 then by actually doing C_2 he does not in the least jeopardise his chances of getting the $\$M$ and he can only gain \$1,000 by transferring himself from *a* to *b*—he cannot transfer himself from *a* to *d*.

The basis for claiming in the first place that experience supports the contention that it is best to do C_1 has disappeared. Past experience is now interpreted in a manner that it no longer provides any grounds for the agent to say 'I had better be careful and do C_1 so as to increase my chances to receive $\$M$'. And it is not because the perfect record of the 'Predictor' is put down as an incredible coincidence that we say that the agent may ignore the 'Predictor'. This was only necessary as long as we were not aware of the possibility of a theory according to which the 'Predictor's' success is indicative only of his proficiency as a diagnostician. Once this is brought to our attention we realise that it is due to the extremely high correlation between T_i and C_i that he manages to be a predictor at all but actually he has no direct access to the final choices.

3 Now we shall look upon the positive reason why on the pain of a contradiction only, could we maintain that the 'Predictor' is skilful *qua* predictor. I argued because that would imply that it is best to do C_1 while the 'well-wisher argument' implies that it is best to do C_2. This is basically different from the 'dominance argument' which has been employed by some authors to support their contention that it is best to do C_2. Cargile sounds somewhat sceptical and is not entirely convinced that the two arguments are fundamentally different. In my original paper I showed that the two arguments must be different by describing what I called Game 2, in which an Observer replaces the Predictor and in which it is clear beyond doubt that it is best to do C_1. Yet the 'dominance argument' applies equally well here as in Game 1 and would force upon us the obviously wrong conclusion that it is best to do C_2, while the 'well-wisher argument' cannot be applied to this game.

It is not enough however, merely to show that the two arguments are entirely different and that the 'dominance argument' *must be* erroneous. It is also very important for the sake of avoiding Cargile's consequent confusion to show *how* the two arguments differ and *why* exactly the 'dominance argument' is wrong. Let me try.

Box II of course may be in one of two states: *E* (empty) and *F* (full). The 'dominance argument' correctly reasons that the agent is better off both relative to *F* and relative to *E* if he takes both boxes and gains the money contained in the first box. It, however, happens to be a fact that to be worse off relative to *F* is preferable than to be better off relative to *E*. If we assume the Predictor to be infallible, then the only two available outcomes are 'worse off *F*' or 'better off *E*' and it is up to the agent to choose between these two. There is nothing in the 'dominance argument' which denies this. Thus, suppose the agent asks himself 'Am I not increasing my chances to find $\$M$ in box II if I restrain myself and refrain from taking box I as well?' There is nothing in the 'dominance argument' which would indicate that he is not increasing his chances of finding money in box II by doing C_1. It supports no more than the contention that should the agent do C_1 then he is to lose the \$1,000 of box I, while doing C_2 will gain him the \$1,000 contained in box I. This in no way excludes the possibility that nevertheless by doing C_1, rather than C_2, that the agent is better off. It is obvious therefore that without violating the 'dominance argument' we may advise the agent to go for the 'worse off *F*' and do C_1.

It is entirely different with the 'well-wisher argument'. To the well-wisher we could put the explicit question: is there any point in taking box II only, rather than both boxes? Does the agent in any way lessen his chances for finding $\$M$ in box II by doing C_2 rather than C_1? We can work out with absolute certainty what his answer would be: there is nothing to gain by doing C_1, the agent cannot in the least increase his chances for obtaining $\$M$ by restraining himself and taking box II only (for his chance is already 100 per cent or in any case will remain 0). And of course, on the assumption that between t_0 and t_1 the state of box II does not undergo any changes, he must be right. We are forced therefore to withdraw the assumption that the predictor is competent, an assumption which leads to an answer conflicting with that of the person who is in a perfect position to judge the issue. From the 'dominance argument' we cannot extract any answer to this question. It therefore cannot force us to revise our assumption that the agent has reason to be concerned that his actual choice may affect the contents of box II.

4 We are in a position now to look at Cargile's objection in detail. He claims that my contention that the agent must follow the advice of the well-wisher ignores the fact that:

> . . . the observer may advise this (*i.e.* to do C_2) for two different reasons: (1) at least you'll get a thousand and (2) you might as well get an extra thousand. If the player subscribes to a theory about the predictor according to which whether it is (1) or (2) depends on his choice, the fact that the observer will always advise C_2 may be irrelevant.

It is indeed true that the competent judge who we know holds the opinion that, in order to maximise his gain, the agent ought to do C_2 may think so either because he sees that there is anyhow no money in box II or because he is already assured that no matter what the agent does, he is going to win the $\$M$ which is in box II, and we do not know which. But if we say that by doing C_2 the agent may bring about, or even just may raise the probability that the observer's deeming C_2 the better choice is because of P_2 rather than P_1, then we are forced to the conclusion that he ought to refrain from C_1. But this is contrary to the views of the one person who is entirely competent to judge this issue and whose opinion is therefore not irrelevant but completely binding.

Or, to put it differently, the agent may raise the question: The well-wisher advises me to do C_2. But by actually doing C_2 am I not raising the probability that he advises me thus because he sees that box II is empty rather than because he sees that there is $\$M$ in it anyhow? But this very question may be put to the well-wisher and we know for sure what his answer would be: No, by doing C_2 rather than C_1 all you do is raise by 100 per cent the probability of winning the \$1,000 contained in box I but affect in no way the probability of there being $\$M$ in box II.

Cargile seems to sense the untenability of his position and tries to turn around in the right direction by saying:

> It might be replied that once the sympathetic observer checks the boxes, it is too late for the predictor to do anything, so the fact that the well-wisher will necessarily advise C_2 cannot be irrelevant.

But he immediately sinks into an unfortunate confusion:

> But this is just to reject the possibility of probabilistic reverse causality which is implied in accepting probabilistic independence. This may be 'reasonable', but then, given this attitude, there is no conflicting strategy, and it is not necessarily true that this attitude is correct.

But of course 'this attitude' is not regarded as '*given*'. To begin with it is suggested that we go along with Newcomb's story and definitely assume that we are confronted with a case of some 'sort of probabilistic reverse causality', *i.e.* that the agent's choice at t_1 influences the Predictor's action at t_0. But after it has been demonstrated to us that ultimately this assumption leads to the conclusion that C_1 maximises the agent's gain, which conflicts with the conclusion following the 'well-wisher argument' that it is rather C_2 which does so, we are forced to withdraw our original assumption. Thus our proof is essentially a ***reductio ad absurdum*** proof. We begin by postulating that the Predictor of our story is competent and demonstrate that the story leads to a contradiction. In order to remedy the situation we reconsider matters realising that one of the assumptions incorporated in our story must have been wrong. This leads us to the conclusion that the wrong assumption was that the Predictor was competent.

5 Benditt and Ross say that there are three major points in my argument:

> First, there is the claim that the well-wisher would wish the player to take both boxes. Second, there is the claim that whatever he wishes the player to do is what is in the player's best interest (this is supposed to be analytic). And third, Schlesinger supposes that what is in the player's best interest is what is rational for him to do.

They are quite correct in saying that I subscribe to the first two points. But of course instead of their third point, I should substitute 'What is *known* to the player to be in his best interest is what is rational (in the sense that it will most promote his best interest) for him to do'. This obviously renders their 'counterexample' entirely irrelevant since in the 'simple game' they have devised the agent does *not* know that the well-wisher who is well-informed about the contents of box *A* and *B* is of the opinion that he should take box *B*. Hence he has nothing to go by but the probability judgments available to him according to which taking *A* is most likely to gain him the money. In our case however the agent knows with absolute certainty that the well-informed onlooker thinks that the agent maximises his gain by taking both boxes. Thus if the agent wants to maximise his gain, he must act in accordance with the former's judgment. Consequently it is wholly pointless to go on arguing as they do that:

> . . . we can and must distinguish between (*a*) what I should do in the sense of what rational policy I should follow and (*b*) what I should do in the sense of what will in fact attain the most desirable result for me.

In a situation like ours where the agent can work out exactly what policy the fully competent judge thinks will bring him the most desirable result, there is no distinction between this policy and the one he ought to follow if he wants to bring about the most desirable result.

6 Contrary to the claims of Benditt and Ross, the fact that the agent knows that in the opinion of the well-informed and intelligent onlooker's judgment it is best to do C_2 is sufficient for the agent to be fully assured that to do C_2 must be the most desirable thing for him to do even though the agent does not know why the well-wisher has this opinion.

Consider the following game: There are two boxes X and Y and a well shuffled pack of ordinary cards. One card is drawn at random and placed in box Y and the other 51 cards in box X. Two dice are rolled and if a double-six occurs the agent is to get a million dollars if he points his finger at the box which does not contain the ace of spades; otherwise he gets a million dollars if he points his finger at the box which does contain the ace of spades. The agent does not know which card has been put in box Y nor the result of rolling the dice and is invited to point a finger at one of the two boxes. The probability that by pointing at box X the agent will get a million dollars is at least $35/36 \times 51/52$ since the probability that the dice showed not a double-six is 35/36 (and in which case he has to point at the box which does contain the ace of spades) and the probability that the ace of spades is in box X is 51/52. In other words the probability that by pointing at box X he will win a million dollars is more than 95 per cent.

Now let us suppose that a friend who is an absolutely perfect well-wisher is allowed either to observe the fall of the dice or to examine the contents of the two boxes but not both. The agent knows this with absolute certainty but has no idea which of the two acts his friend was allowed to do. Let us also suppose that the friend is permitted to communicate with the agent but no more than to advise him at which box to point. Consider the case in which the friend says to the agent 'Point at Y'. It is clear that he may have given this advice for two different reasons. Either because he saw that the improbable has happened and both dice showed a six in which case he wins $\$M$ if he succeeds in pointing at the box not containing the ace of spades and there is a probability of 51/52 that the ace of spades is not in Y, or because he has seen the contents of Y and knows that the single card in it turned out to be the ace of spades in which case there is a probability of 35/36 that he will gain $\$M$ by pointing at it since the probability of not having a double-six is 35/36 which is the probability that he wins $\$M$ by pointing at the box that does contain the ace of spades. There is not a shadow of doubt that it is rational for the agent to point at Y. It matters not in the least that he has no idea for which of the two possible reasons his friend has advised him to point at Y, the mere fact that he advised him thus suffices to ensure that by pointing at Y the agent maximises his chances to win the money. Thus it is quite pointless for Benditt and Ross to say:

> ... in the Newcomb's game ... what choice one's friend would have one to make does not provide any information about what the situation is.

The agent need not know 'what the situation is', *i.e.* on what information does his friend base his opinion that it is best to do C_2. It is enough for him to know that the final verdict of a perfectly competent judge is that by doing C_1 he can gain absolutely nothing. Doing C_2 must therefore be his best choice.

GEORGE SCHLESINGER
The University of North Carolina at Chapel Hill

Brit. J. Phil. Sci. 27 (1976), 275–287 *Printed in Great Britain*

Review Article

RETHINKING TRANSFORMATIONAL LINGUISTICS*

1 *Introduction.*
2 *The Problem Background.*
3 *The 'Empirical Basis': the Problem of 'Intuitions' as Linguistic Data.*
4 *The 'Psychological Reality' of Linguistic Constructs.*
5 *Conclusion.*

1 INTRODUCTION

Transformational linguistics has, for the past twenty years and particularly in America, been the major focus for linguistic research.[1] Its doctrines have, in Lakatosian terms,[2] provided both a hard-core of assumptions about the nature of language and a research strategy (heuristic) in terms of which these very abstract assumptions have been articulated into increasingly more complex empirical theories about language. Nevertheless, while there is still considerable consensus about the major transformational doctrines, there has been, increasingly in recent years, rather heated controversy about some relatively minor, but by no means peripheral, transformationalist assumptions. In this paper, I will discuss some of the methodologically more interesting aspects of this controversy. I will do this primarily by way of some comments on two recent foundational studies in this field: Geoffrey Sampson's *The Form of Language* and Bruce Derwing's *Transformational Grammar as a Theory of Language Acquisition*. And I will, moreover, try to sort the rather important wheat from the occasionally annoying chaff in these works by means of some elementary methodological observations.

This paper should be understood, then, both as a critical commentary on some recent work in this field, and as a necessarily sketchy prolegomenon to a more thorough discussion of the issues touched upon here. The paper is divided into four parts. In 2, I set the stage for what follows by means of a brief discussion of some of the assumptions common to most transformational research. In 3–4 I survey two specific areas of disagreement in present day transformationalism—in particular, some problems concerning the nature of linguistic data, and the question of the 'psychological reality' of theoretical constructs in transformational linguistics. In 5 I offer a brief conclusion and make some rather more general remarks about Sampson's and Derwing's books.

---

* Review of B. Derwing [1973]: *Transformational Grammar as a Theory of Language Acquisition*. Cambridge University Press. £3.25. Pp. 346; and G. Sampson [1975]: *The Form of Language*. London: Weidenfeld and Nicholson. £3.50. Pp. 236. My thanks are due to John Worrall and Gregory Currie for helpful comments on previous drafts of this paper. Geoffrey Sampson has also commented helpfully on some of the topics I have discussed here, though he has not seen any version of the paper itself.

[1] *Cf.* Chomsky [1965], section I for a discussion of the research aims and methods of transformational linguistics.

[2] *Cf.* Lakatos [1970] for his 'methodology of scientific research programmes'.

2 THE PROBLEM BACKGROUND

As I have already indicated, there are several assumptions about the nature of language and of the linguistic enterprise about which there is considerable consensus among transformationalists. It is these common assumptions which distinguish them from their 'structuralist' predecessors and their present-day opponents. Thus, transformational linguists are concerned primarily with two problems: (1) writing grammars of individual human languages, and (2) providing an explanation of human language learning.

For the transformationalist, a grammar is a formal system—of a kind quite similar, in fact, to the formal systems of mathematical logic.[1] Basically, a grammar serves to divide the set of possible strings of words of a language into 'grammatical' and 'ungrammatical' subsets of strings. And it is interpreted as predicting, given certain auxiliary assumptions, that users of the language in question will employ the grammatical but not the ungrammatical strings. In this way, then, a grammar of a language serves—in conjunction with auxiliary assumptions—to explain why users of that language employ certain strings of words, but not others. Furthermore, transformational linguists characteristically interpret a grammar in a 'realistic' way. Thus, they do not consider it merely as an instrument for the prediction of linguistic behaviour. Rather, the grammar devised by the linguist for some language is taken to be 'psychologically real' for its users. That is, transformationalists interpret a grammar as representing what it is that users of a language must 'know' about that language in order to employ it as they do. Thus, a grammar explains linguistic behaviour ('performance') by pointing to a psychological reality which underlies that behaviour (the user's 'competence' to use his language). A grammar is, in effect, a description of this competence.[2] Thus, the problem of writing grammars for various human languages becomes, for the transformationalist, the problem of accounting for what it is that users of those languages 'know' about them.

This way of formulating this first problem immediately suggests the second important problem for transformational linguists: That is, if users of a language 'know' something about this language, then how do they acquire this knowledge? Transformationalists have characteristically answered this question in a 'rationalist' manner. Thus, on their account, language learners acquire 'knowledge'—*e.g.* about regularities in the linguistic data to which they are exposed during the learning process—not by 'induction' from the data (as many non-transformational linguists assume), but because they are innately predisposed to discover just those kinds of regularities.[3] Thus, this innate predisposition provides a kind of 'schema' which guides the language learner in constructing the 'psychologically real' grammar that comes to underlie his linguistic behaviour.

---

[1] My comparison of grammars and formal systems is by no means gratuitous. Chomsky's realisation of the fruitfulness of just such an analogy had a major heuristic influence on the development of transformational linguistics. *Cf.* Chomsky [1959], the touchstone paper for this development, and my [1976] for some of the evidence of the fruitfulness of this analogy.

[2] A philosophically sound way of formulating this thesis has proved to be difficult to discover. Thus, Chomsky's claim that a language user 'knows' the rules of the grammar of his language has been attacked, *e.g.*, by Harman in his [1967].

[3] That language learners have an innate predisposition to acquire linguistic knowledge of a certain kind has also proved to be a difficult thesis to articulate in a philosophically unobjectionable way. *Cf.*, for instance, Goodman's [1967] for a criticism of Chomsky's position on this issue.

It is this rationalist assumption about language learning and the realist assumption about the nature of linguistic competence that form a hard-core of assumptions around which research in transformational linguistics has been oriented. In addition to these general assumptions, certain more specific assumptions about the nature of language and of the linguistic enterprise have recently become focal points for controversy among transformationalists. In the next two sections I examine more closely two of these disputed points and offer some comments on the debates to which they have given rise.

3 THE 'EMPIRICAL BASIS': THE PROBLEM OF 'INTUITIONS' AS LINGUISTIC DATA

Many, but not all, transformational linguists have felt that some of the things that language users 'know' about their language are reflected in certain 'intuitions' that users have about that language. These linguists hold, for instance, that users of English 'know' intuitively that (*i*) below is related to (*ii*) in a way that neither is to (*iii*), and that these three sentences are grammatical, while (*iv*) is not.

(*i*) *The boy loves the girl.*
(*ii*) *The girl is loved by the boy.*
(*iii*) *The train was late.*
(*iv*) *Scat fireplace mat nicely.*

Furthermore, these linguists feel that these sorts of intuitions constitute a crucial kind of data for the construction of grammars. Other linguists, on the contrary, feel that the only data relevant to the construction of grammars is the set of sentences which users are actually observed to employ. Linguists who take this second position characteristically do so for one of two reasons: (*a*) Because, while taking a 'psychologically realistic' view of grammars, they nevertheless feel that intuitions are an insufficiently reliable guide to what it is that users can be said to know about their language. Or (*b*) because they feel that users cannot, in any reasonable sense, be said to know *anything* about their language.

Linguists who, on the other hand, defend the relevance of intuitions as data have characteristically done so on a number of grounds. (*A*) Because users are held to have intuitions not only about the grammaticality and ungrammaticality of certain strings of words, but also about certain structural relationships between strings of words. Thus, taking these intuitions into account effectively expands the range of data for which a grammar has to account. For, if we count as data only the strings which users are actually observed to employ, this provides evidence only about the grammatical sentences of the language, and none whatsoever either about the ungrammaticality of certain strings or about relationships between strings. Thus, on this argument, intuitions are relevant to the construction of a grammar on general methodological grounds: Grammars are more falsifiable relative to the set of intuitive data than they are relative just to the data of observed language use.

(*B*) Linguists generally agree that grammars ought not to be required to mark as 'grammatical' *all* the strings of words which users are actually observed to employ. For, for instance, users often interrupt themselves in mid-sentence or, in particularly complicated utterances, lose track of the intended structure of the sentence. And, the argument goes, we certainly don't want a grammar to have to account for these kinds of phenomena. That is, most linguists would agree on the need to draw some distinction between actually observed linguistic behaviour and the 'idealised' linguistic behaviour which a grammar is intended to explain. Thus, linguists who defend the relevance and importance of intuitions as data have often claimed that it is *only* the user's intuitive knowledge that a particular *observed* string of words is *un*grammatical that justifies the linguist's decision to exclude this string from the corpus of data for which a grammar is to account. That is, it has often been argued that it is only the use of intuitive data that justifies the linguist's distinction between actual and idealised linguistic behaviour.

Now, as long ago as 1957, Chomsky recognised that, due to a variety of essentially non-linguistic (or at least non-syntactic) factors, such linguistic intuitions could not always be trusted to be completely accurate reports of the user's underlying knowledge of his language. Thus, we are sure to think of (*v*) below as 'odd'. Is this because (*v*) is 'ungrammatical'? Or is it, rather, because (*v*) is 'non-sensical', though perfectly grammatical?

(*v*) *Colourless green ideas sleep furiously*

Of course, only a fully developed and independently supported theory of those factors which interfere with the expression of genuine lingustic intuitions would enable us to decide, in troublesome cases like this, just what the user's judgments were a report of. Nevertheless, Chomsky felt that linguistic research could proceed, even in the absence of such a theory, since there were a great many *prima facie* 'clear cases' in which genuine linguistic intuitions could be assumed to have been reliably reported.[1] And these 'clear cases' could, he felt, serve as a sufficient data base for the construction of grammars. For a number of years, this assumption was, it seems, tacitly justified by the successful progress of such a grammar-constructing enterprise. In recent years, however, it has seemed that precisely the cases which have been crucial to deciding between different grammatical analyses have not been clear, but very unclear indeed, intuitions about which often vary even from individual to individual.[2] This fact has, according to some linguists, arrested progress in linguistic research, substituting for it instead an often irrational-seeming and uniformly inconclusive debate about the empirical status of disputed intuitive judgments.[3] This, in turn, has led an increasing number of linguists to argue that intuitions should, given their unreliability, be altogether abandoned as a source of linguistic evidence.

Sampson is one such linguist and the arguments he gives for dispensing with the use of intuitive data are among the clearest I have encountered. Sampson's basic strategy in arguing against the use of intuitive data is this: First, he considers the major argument in favour of using such data—i.e. (point (B) above) that only users' intuitions enable the linguist to distinguish the strings of words which users are actually observed to employ from the grammatical sentences for which a grammar is intended to account. Then he tries to show how exactly this same distinction can be drawn on much more general methodological grounds. That is, Sampson tries to show how linguists don't need to use intuitive data in order to justify their idealisations away from actually observed linguistic behaviour.

---

[1] *Cf.* Chomsky [1957], 13–14.

[2] *Cf.* Fillmore [1972].

[3] *Cf.* Botha [1973].

Let me now trace his argument more or less point by point.[1] Sampson first considers the following problem. He has decided to eschew the use of intuitive data and is now forced to fall back on observations of strings of words which users of a language actually employ as the only source of data for the construction of a grammar. But this decision raises the problem of 'negative instances'. That is, we know that any sentence that we have actually observed users employing will have to be accounted for by our grammar, but what of those possible strings of words which we have not observed? Perhaps if we observed for a longer time, we might observe some of them. Sampson then asks whether, in view of this 'asymmetry', our best move might not be to construct a grammar which treats all possible strings of words as grammatical. But this, he recognises, would surely be counterintuitive, for some strings of words are 'known' to be ungrammatical. Surely, in this case, we must rely on users' intuitions that this is so in order to exclude such sentences. This is Sampson's first 'impasse'. He avoids it by pointing out that in general and as a matter of methodological practice, we require any scientific theory to be as strong as possible, and that we are thus warranted in constructing a grammar which defines as 'grammatical' the narrowest possible range of strings consistent with those which have actually been observed. Thus, we don't need users' intuitions to tell us that not every string of words ought to be marked as grammatical by our grammar.

While Sampson's solution to this first impasse is a clever one, it must be noted that something is lost in solving the problem of 'negative instances' by means of a purely methodological convention. For, while methodology tells us only that some, indeterminate set of strings of words need not be accounted for by our grammar, our intuitions can, ideally, tell us precisely which strings must be excluded as data for which the grammar must account. Thus, in shifting the problem in the way he has, Sampson has radically narrowed the data base for the construction of grammars. Of course, since he—not without considerable justice—considers intuitions as an extremely unreliable kind of data, Sampson may feel—again with some justice—that such a narrowing is necessary and, in fact, appropriate. He may feel that his data base, while narrower than one which includes intuitions, is, correspondingly, much more secure.

In any event, Sampson's solution to this first problem leads him directly to a second one. For this methodological directive (of maximising the strength of grammars) seems to imply that we should construct grammars which mark as 'grammatical' all and only the strings of words which we have actually observed. But, as he points out, some strings which we happen not to have observed will be known, intuitively, to be grammatical. This is his second 'impasse'. What reason, other than intuition, can we have for including such sentences in the data which our grammar must explain—particularly since such inclusion will 'weaken' our grammar? His answer to this question is simple: We can include such sentences because, in order to exclude them, we would have to complicate the rules of our grammar to a degree unwarranted by the increase in strength that results. And, he points out, it is a general methodological principle that bids us to construct the simplest possible theories consistent with the observed data. Here too, then, according to Sampson, we do not need intuitive data in order to justify our decision about what data a grammar can be appropriately required to explain.

---

[1] In what follows in this section I summarise and occasionally expand the argument given in Sampson [1975], section IV.

Sampson's solution to this second 'impasse' is again a clever one but is, equally, subject to certain difficulties. There are two main problems here. First, this solution hinges quite crucially on the idea of simplicity involved. Now, within the central tradition of transformational research, this idea does have a fairly well-defined sense. A 'second-level' theory about the kinds of possible grammars defines, in effect, a canonical language within which grammars of human languages are written. The 'simplest' grammar of a given language is then, very roughly, just that grammar, written in the canonical language, that uses the fewest symbols of that canonical language. Unfortunately for the viability of his position here, Sampson rejects this notion of simplicity and proposes to substitute for it a pretheoretical, philosophically 'common-sense' notion. This decision, however, weakens his position on this question considerably since, as philosophers are well aware, presystematic notions of simplicity have generally proved to be controversial, if not downright intractable.

The second difficulty with Sampson's solution to his second impasse lies with the difficulties involved in balancing decreases in strength against increases in simplicity. Sampson is quite correct in noting that increasing the simplicity of a grammar often means compromising its strength. But, since Sampson seems to feel that the decreases in strength involved are often 'tiny' in comparison with the corresponding increases in simplicity, the problem of balancing these increases and decreases may seem to him to be a simple one. Nevertheless, Sampson is either wrong or insufficiently clear when he claims that the decreases in strength involved are 'tiny'. For, in order to maximise the simplicity of a grammar we may be obliged to consider as grammatical a whole range of unobserved sentence *types*, each one of which might have a potentially infinite number of *tokens* or exemplars. And the decrease in strength that results from making this kind of decision is surely not a tiny but, rather, an infinite one. It is a question left open in Sampson's presentation how infinite decreases in strength are to be balanced by increases in simplicity. This is a problem which will have to be solved before Sampson's methodological solution to his second impasse can be considered an acceptable substitute for the use of intuitive data. (Notice, by the way, that this problem doesn't even arise when we permit the use of intuitive data. For, since users have intuitions both about grammaticality and ungrammaticality, every grammar of a language is equally strong. Thus, a 'simpler' grammar of some language might predict that a much larger set of sentences are grammatical than some more 'complex' grammar does. Nevertheless, these grammars are equally falsifiable. Thus, where the more complex grammar could be falsified by the observation of some sentence marked as 'ungrammatical' by it (though, perhaps, as 'grammatical' by the simpler grammar), the simpler grammar could, in turn, be falsified by the intuition that some sentence marked 'grammatical' by it (though not, perhaps, by the more complex grammar) is, in fact, 'ungrammatical').

So far, Sampson has assumed that a grammar should mark as grammatical *all* the sentences which we have actually observed and has worried through the problem of sentences which we haven't observed. Next he addresses a far more difficult problem—one which has been the last and best defence of those who argue for the use of intuitive data. This is the problem created by the fact that some observed sentences are known to be ungrammatical. Aside from invoking our intuition that this is so, how can we exclude such sentences from the corpus of data which our grammar is required to explain? Sampson's solution to this problem is to point out that a grammar can be construed as predicting that a string that it marks as ungrammatical will not be employed by users if and only if 'everything else is equal'. Thus, a grammar might predict that (*vi*) but not (*vii*) is grammatical.

(*vi*) *If the phone rings, answer it.*
(*vii*) *If the phone.*

What if we observe (*vii*)? Does this falsify our grammar and so require us to construct a new one which does mark (*vii*) as grammatical? Not at all, says Sampson. We can simply point out that, in this case, everything else wasn't equal; that, for instance, the user who employed (*vii*) meant to employ (*vi*) but was, for instance, interrupted by some other task. Ideally, we might even be able to adduce independent evidence for this 'saving auxiliary hypothesis' and thus justify our decision not to regard the observation of (*vii*) as a refutation of our grammar. Even more ideally, elaborate non-linguistic hypotheses would supplement our grammar and thus permit us to exclude, on a principled basis, many of those observed sentences which are known, intuitively, to be ungrammatical. Here too, then, the use of intuitive data seems to be unnecessary.

This solution may seem to be parasitic on the tacit guidance of intuitions in the sense that it may seem that we would invoke such auxiliary hypotheses only when our intuition told us that a particular sentence, while observed, was, nonetheless, ungrammatical. This point, however, has no real force against Sampson's solution. For Sampson freely admits that many of our decisions about the possible grammaticality of observed sentences will be *motivated* (in the psychological sense) by our intuitions about them. What he is claiming is that our decision to exclude an observed sentence as ungrammatical need and must not be *justified* by appeal to intuitions. His claim is that such justification is better provided by the use of explicitly formulated and independently testable auxiliary theories.

By the lines of reasoning just sketched, Sampson finally arrives at his general conclusion: Intuitions are not necessary in order to justify either our decision to consider certain sentences as grammatical even though we have never observed them or our decision to consider some sentences actually employed by users as ungrammatical; such decisions can be justified either on methodological grounds or by means of independently supported auxiliary theories.

This solution is clever even if, as I have pointed out, not entirely without its weaknesses of detail. Moreover, it certainly seems to solve the problem that Sampson wanted to solve: *i.e.* how to show that linguistic intuitions simply aren't necessary in order to distinguish between the linguistic behaviour of users for which a grammar ought to be required to account and that for which it needn't account.

But this solution does not, frankly, seem to solve another and more pressing problem. Recall that I mentioned that many linguists were increasingly suspicious of intuitive data largely because such data were not sufficiently reliable to decide between different proposed grammars of the same language. The problem, roughly, is this: One grammar G_1 predicts that a sentence S is grammatical, while another grammar G_2 predicts that S is ungrammatical. But our intuitions about S are unreliable and, so, we cannot decide between G_1 and G_2. (Typically, the unreliability is occasioned by the fact that one linguist thinks that S is grammatical, while another thinks that S is ungrammatical.) Now, on Sampson's account we should in such a situation simply ask whether or not S has been observed. (Notice that if S *hasn't* been observed our problem remains. For a failure to observe S is consistent both with its grammaticality—for many grammatical sentences are likely to remain unobserved—and with its ungrammaticality.) But, does finding that S has been observed solve the problem—*i.e.* by refuting G_2, which predicted that S is ungrammatical? Not at all. For S might have been observed for reasons that have nothing to do with its grammaticality, as, for instance, (*vii*) above was. Of course, Sampson can reply that, as in the case of (*vii*), we may have strong, independent reasons for supposing that S was observed in spite of its ungrammaticality, rather than because of its grammaticality and thus, to continue this line, that observing S does not refute G_2 but, rather, G_1. Such an argument is certainly plausible when it is a sentence like (*vii*) that is at issue, but what, for instance, of (*viii*) below, recently much discussed as a crucial example whose putative ungrammaticality favours one analysis and disfavours another?

(*viii*) *Harry reminds me of himself.*

Well, we may have some reasons for supposing that (*viii*) is not ungrammatical and thus does not refute the grammar that marks it as grammatical,[1] but it seems quite likely that such reasons are neither clear and specific enough nor well enough justified independently to decide the issue one way or another.

This conclusion points to what is really the moral of my story here: What we need in order to be able to decide issues like these are very strong and independently supported theories about, in Sampson's terms, what makes apparently ungrammatical sentences usable and vice versa. It is only if we have such theories that Sampson's data base for linguistics (*i.e.* observed sentences) will be any more reliable or objective than the use of intuitions. For, without such theories, present-day squabbling about the reliability of intuitions will simply be replaced by squabbling about whether a particular sentence was observed because it is grammatical or in spite of its ungrammaticality.

Of course, Sampson may feel that it is easier and more productive to squabble about fully articulated auxiliary theories than it is to argue about intuitions. For, when the validity of an auxiliary theory is disputed we can, Sampson might say, always try to articulate the till-then tacit assumptions that lie behind it. The debate about such auxiliary theories, Sampson might argue, can be conducted in a more objective manner than a debate about the validity of intuitions. This line of argument is a promising one, but it is not clear to me that it is a decisive argument against the use of intuitive data. It is so only if we feel, as, alas, many linguists seem to, that intuitions about language are somehow intrinsically incorrigible. But I hardly see that linguists, whatever their *opinions* on this subject, are *forced* to hold such a position. It seems quite reasonable to hold, for instance (and this is roughly Chomsky's position), that the intuitive judgments of language users are only reflections of and not direct reports of their tacit 'knowledge' about their language. We need only recognise, that is, that many factors other than the users' knowledge are mixed up with that knowledge in reports of his intuitive judgments. And, in fact, the kinds of auxiliary theories that Sampson will need if his proposal is to be viable can just as easily, perhaps, be interpreted as theories about the non-linguistic factors which influence and distort intuitive judgments.

---

[1] As Bar-Hillel suggests we do in his [1971]

To see how this might work, consider the following situation: Linguist A proposes a grammar G_1 for some language L. G_1 marks the string S as grammatical and this accords with A's intuitions about S. Linguist B, on the other hand, proposes a grammar G_2 for L. G_2 marks S as ungrammatical and this accords with B's intuitions about S. Now, A and B *could* proceed to *squabble* about the validity of their intuitions. This is likely to be pretty unproductive, as Sampson rightly contends. On the other hand, A and B could, more productively, try to *criticise* each others' intuitions by means of explicitly formulated auxiliary theories. For instance, A could say to B: 'Your intuition that S is ungrammatical is mistaken. S may be "unacceptable" to you but it is, nonetheless, fully grammatical. Its unacceptability is due, moreover, not to its ungrammaticality, but (for instance) to its "semantic oddity". In fact, S is semantically odd in just the same way as many other strings about whose grammaticality you and I agree. Thus, your intuition about S is an intuition of its semantic oddity, not of its ungrammaticality. Furthermore, since we agree that other strings which are semantically odd in the same way as S *are* grammatical, you ought to agree with me that S is grammatical too.' Of course, B may have a counter-proposal here and it may, in practice, be as difficult to resolve a debate couched in this form as it is to resolve squabbles about intuitions. Nevertheless, a debate of this kind seems to me to be fully as 'objective' and 'productive' as one between two Sampsonian linguists who are trying to decide whether S was *observed* because of its grammaticality or in spite of its ungrammaticality.

Thus, on this account, Sampsonian auxiliary theories would serve as 'observational theories' in terms of which linguists could 'interpret' statements about intuitions in order to tap the genuine 'knowledge' that lies behind them. And, in this case, intuitions would be just as reliable a data base as observed sentences. The basic point to be made here is a simple one: There is nothing more objective about sentences which are 'observed' by means of auxiliary observational theories than there is about intuitions which are 'interpreted' in something like the same way. Adopting Sampson's proposal does not, in the absence of such theories, contribute to the clarification of certain crucial debates in linguistics. And, further, if such theories were available, then it would not be clear that we would *have* to adopt his proposal.

4 THE 'PSYCHOLOGICAL REALITY' OF LINGUISTIC CONSTRUCTS

As I pointed out in 2 above, transformational linguists have characteristically held that grammars which they construct for various languages are 'psychologically real' in the (relatively indeterminate) sense of somehow modelling or reflecting the tacit 'knowledge' of users of those languages. Let me call this position the Weak Thesis about the 'psychological reality' of grammars. This Weak Thesis—while itself not nearly specific enough to be considered an empirical theory—has, nevertheless, given rise to a Strong Thesis about the psychological reality of grammars which can be tested empirically. According to this Strong Thesis, the grammars constructed by linguists are 'isomorphic' with the tacit knowledge that users employ in producing and understanding utterances.[1] In fact, the viability of this Strong Thesis has recently been challenged on a number of grounds.

Thus, in his book, Derwing points out that the Strong Thesis is, quite simply, refuted by much experimental psycholinguistic evidence.[2, 3] He draws two conclusions from this—one I think wrong and the other rather more to the point. First, he concludes that these negative psycholinguistic results suggest that not only the Strong but also the Weak Thesis must be abandoned. But this is surely wrong. For all the Weak Thesis implies is that there is some *unspecified* way in which grammars constructed by linguists are employed by language users. And this thesis can surely be maintained even in the face of a refutation of the Strong Thesis. Of course, proponents of the Weak Thesis will have now, given such a refutation, to find some new way of strengthening it into an empirically testable hypothesis—for as it stands, it is not itself so testable. And they will have to do this in some non-*ad hoc* and content-increasing way. But, if they *can* do this, there is no reason for linguists to abandon the Weak Thesis.

Perhaps the best way to look at this situation is as follows: Transformational linguistics constitutes not a theory, but a research programme (in Lakatos's sense) about the language user's competence. The Weak Thesis forms the hard core of that programme. Its heuristic directs linguists to elaborate more and more complex models of the interaction between an internalised grammar and other factors in order to explain how users actually process sentences. The first such model—the simplest imaginable—was the 'isomorphic' one. It gave rise to a testable theory—the Strong Thesis that Derwing rightly contends has been refuted. But only this first theory and not the research programme as a whole needs, on this ground, to be abandoned. Thus, it now still remains to linguists to try to elaborate a somewhat more complex model of the interaction between a grammar and other factors in processing. In fact, some such suggestions for elaborating the model are already at hand.[4] There are, in any event, only two circumstances under which linguists could be expected to abandon the Weak Thesis itself. First, if, after repeated attempts to produce an adequate account of language processing, no such account incorporating the Weak Thesis had been found. And second, if some account of language processing which did *not* incorporate the Weak Thesis *did* provide an adequate account of language processing.[5] Derwing has not suggested that either of these conditions obtain, so his conclusion that linguists ought to reject the Weak Thesis cannot be sustained.

---

[1] On the thesis that users' and linguists' grammars are 'isomorphic', *cf.* Levelt [1974], Vol. 3, section III.

[2] This view is by now generally accepted. *Cf.* Fodor, Bever, and Garrett [1974], sections VI–VII for a report of the research which led to this conclusion.

[3] This is the gist of Derwing's arguments in his [1973], section VIII.1.

[4] *Cf.* Fodor, Bever, and Garrett, *loc. cit.*, for some such suggestions.

[5] *Cf.*, however, Levelt, *loc. cit.*, for a discussion of some promising alternative models of language processing.

Derwing, however, draws a second conclusion from the refutation of the Strong Thesis with which I am in considerably more sympathy. Thus, he points out that transformational linguists have adopted, as a heuristic principle, the policy of constructing maximally simple grammars. This, he notes, seems to be a hangover from pretransformationalist 'structuralist' days, when linguists were interested in providing the simplest possible classification of the set of grammatical sentences of a language. And, while it was perfectly legitimate for structuralists to proceed in this way—in particular because they made no additional claims about the psychological reality of their grammars—this may not be so for the transformationalists—who do make such claims. For, Derwing argues, the goals of maximum simplicity and psychological fidelity may not be consistent. Thus, he insists, it is an empirical question whether users' (*i.e.* psychologically real) grammars are maximally simple. Further, there is some evidence that they are not maximally simple. (In particular, the refutation of the Strong Thesis suggests just this.) But, he concludes, if this is the case, then what is the point of adopting the principle of maximising simplicity in writing grammars that are supposed to be psychologically real? *Prima facie*, are these two goals inconsistent with one another?

This conclusion seems to be rather more to the point than Derwing's first, rather more dramatic conclusion. For it surely is the case, *as a matter of actual practice*, that linguists do often make rather naive assumptions about the relation between simplicity and psychological fidelity. Nevertheless, it would be a mistake to conclude from this that the goals of maximising simplicity and achieving psychological fidelity are intrinsically inconsistent. For these goals may perfectly well be consistent under some methodologically more sophisticated interpretation of their relationship. There are, in fact, two ways in which a more sophisticated view of this situation could vitiate the *prima facie* inconsistency to which Derwing has drawn attention.

First, we need to recall that a grammar is 'simple' only relative to some canonical language in which grammars are written. One way of harmonising the goals of simplicity and psychological fidelity would, thus, be to restructure the canonical language of grammars so that this canonical language better reflected what we know about how users actually process their language. In this case, we could reasonably expect that the 'simplest' grammar of any particular language, being the one that deviates least from the format provided by the now psychologically more realistic canonical language, would itself be psychologically realistic. Of course, since linguists have not yet made any systematic attempts to so restructure the canonical language of grammars, Derwing's criticism is not without its force. But it should be understood, I think, as rather more a criticism of current linguistic practice than of the transformationalist position in the abstract.

A second way in which the *prima facie* inconsistency between these two goals can be vitiated is the following. Thus, it is perfectly possible to interpret the heuristic principle of maximising simplicity as a 'short-cut' heuristic principle. It might, then, be interpreted as saying something like the following: 'It is reasonable to assume that psychologically real grammars are *fairly* simple. In particular, it is much more reasonable to assume that they are maximally complex, in which case it would hardly be possible to write a grammar at all. Furthermore, since we can't know *a priori* how the user's grammar diverges from the maximally simple one, we might as well assume that the user's grammar is maximally simple. This assumption can, in any event, always be tested in particular cases against the psycholinguistic data. If this assumption is, in some particular case, refuted, we can modify our grammar (and perhaps our second-level theory of grammar) accordingly.' Now, used in this way, the principle of maximising simplicity seems utterly inoffensive and certainly not inconsistent with other goals. One can, then, only join Derwing in being disheartened that linguists have almost universally used this principle *not* as a helpful short-cut, but as an infallible guide to the construction of psychologically real grammars. Until linguists get a little clearer about the nature of this principle, transformational research practice *is* going to be inconsistent with transformational research goals. Nevertheless, as before, we must guard against sharing Derwing's unwarrantedly pessimistic conclusions: the present inconsistency is not intrinsic to the transformational enterprise; it stems primarily from the insufficiently sophisticated state of transformational research practice.

5 CONCLUSION

What I have tried to do in this paper is to show how some elementary methodological observations might contribute to the clarification of some currently controversial questions in transformational linguistics. If I have succeeded in interesting other methodologists in these questions they could hardly do better than to read for themselves Sampson's well-written and closely argued book. Reading Derwing's book, which I have also discussed, will require more patience and self-restraint. His frequent forays into methodology of science proper are marred by his eclectic use of incompatible positions—at one point Popper is invoked in defence of a patently inductivist methodology! His book is also raucously polemical. But, although it has many flaws, it contains some good arguments against received transformationalist wisdom.

F. B. D'AGOSTINO
London School of Economics

REFERENCES

BAR-HILLEL, Y. [1971]: 'Out of the Pragmatic Wastebasket', *Linguistic Inquiry*, 3, pp. 401–407.

BOTHA, R. [1973]: *The Justification of Linguistic Hypotheses.*

CHOMSKY, N. [1957]: *Syntactic Structures.*

CHOMSKY, N. [1959]: 'On Certain Formal Properties of Grammars', *Information and Control*, 1, pp. 91–112.

CHOMSKY, N. [1965]: *Aspects of the Theory of Syntax.*

D'AGOSTINO, F. [1976]: 'The Relevance of Mathematical Linguistics to Empirical Linguistics', *York Papers in Linguistics. Forthcoming.*

DERWING, B. [1973]: *Transformational Grammar as a Theory of Language Acquisition.*

FILLMORE, C. [1972]: 'On Generativity', in S. Peters (*ed.*), *The Goals of Linguistic Theory*, pp. 1–19.

FODOR, J., BEVER, T. and GARRETT, M. [1974]: *The Psychology of Language.*

GOODMAN, N. [1967]: 'The Epistemological Argument', *Synthese*, 17, pp. 24–28.

HARMAN, G. [1967]: 'Psychological Aspects of the Theory of Syntax', *Journal of Philosophy*, 64, pp. 75–87.

LAKATOS, I. [1970]: 'Falsification and the Methodology of Scientific Research Programs', in I. Lakatos and A. Musgrave (*eds.*), *Criticism and the Growth of Knowledge*, pp. 91–196.

LEVELT, W. [1974]: *Formal Grammars in Linguistics and Psycholinguistics*.

SAMPSON, G. [1975]: *The Form of Language*.

Reviews

SWINBURNE, R. [1973]: *An Introduction to Confirmation Theory.* London: Methuen. Pp. 218+vi; £3.90.

In writing this review it is necessary to mention some of my own published work but only two recent papers are fully cited: they contain numerous references.[1] Also my own terminology will be used so it will be useful to start with a few definitions.

A *psychological* probability (in my terminology) is one depending on a judgment made without necessarily using mature consideration. A *subjective* (or multi-subjective) probability is based on mature judgment and is part of a body of judgments which 'you' (singular or plural) have made some effort to make coherent or consistent with the usual axioms of probability. Some subjectivists insist on complete consistency, but in any case the purpose of the theory is to help you in your effort towards consistency. It would be more fully descriptive to call subjective probabilities 'logico-subjective', or 'logico-personal', or 'subjective but (roughly) coherent', but 'subjective' is the term in common use among statisticians and is short. The ideal subjective probabilities that would be adopted by a perfectly rational man are called logical probabilities (an expression used in a different sense by Popper) or, following F. Y. Edgeworth and Bertrand Russell, *credibilities.* While recognising that subjective probabilities are more operational than credibilities, I regard it as *mentally healthy* to think of one's subjective probabilities as one's best estimates of credibilities, whether credibilities exist or not. A *physical* probability is what Harold Jeffreys once called an *intrinsic probability,* and which Popper called a *propensity.* A *Bayesian* is a statistician who makes explicit use of subjective probabilities or credibilities. If Swinburne were a statistician he would be a Bayesian.

He says (p. 22) that a subjective theory does not seem to him to give a correct account of any sort of probability statement. He must be referring to a theory of psychological probability because my use of the subjective theory is almost identical with his use of 'epistemic' probability, which is his name for credibility. He says (p. 99), 'The purpose of this book is . . . to draw out [from Swinburne?] the criteria of epistemic probability used by men of our culture today'. Thus he must be concerned with multisubjective probabilities as estimates of credibilities, and he uses his own judgments (influenced of course by those of others) to estimate the multisubjective ones. So his dismissal of the theories of Ramsey, de Finetti, and Savage on page 24, and perhaps the ignoring of my theory, must be based on a semantic confusion. We were not concerned with psychological probability but with subjective probability, just as Swinburne is when you come right down to it.

While on the matter of words, it may be noted that Swinburne calls an unconditional or absolute probability an intrinsic one, thus conflicting with Jeffreys's occasional usage. Also, instead of the well established, though imperfect technical term, *likelihood*, he uses *accuracy* which is even worse because it does not even have the merit of being well established in this sense.

---

[1] Good [1975*a*] and [1975*b*].

Swinburne says (p. 3) he put 'confirmation' in the title of his book because this was Carnap's [bad] term for credibility, but Swinburne uses 'confirmation' in its linguistically *correct* sense. Perhaps the title should have been 'An Introduction to the Theories of Credibility and Corroboration.' Note that corroboration can be best explicated by the excellent technical term 'weight of evidence' which was introduced by C. S. Peirce in 1878 and used by me in about thirty-five publications. It is the logarithm of the 'Bayes factor'. The book contains no reference to this concept.

The book contains a number of closely-knit arguments, and some inconsistencies, but I naturally agree with the conclusions that I have myself published (perhaps sometimes also in ignorance of previous literature). The literature is vast and I am the only living person who has read every word written by I. J. Good, so I cannot blame Swinburne for not having read many of them, but his omission makes the book historically misleading, especially for new students. To clarify some historical issues I shall mention a few anticipations. Based on this sample it seems probable that there are many others.

The device of attaining a simple theory of partially ordered probability, using the usual *numerical* axioms, as proposed on page 38, is identical with my 'black box' approach (1947–50, 1951–2, 1959, 1960–2). A succinct but fairly complete answer to Ayer's question (p. 27) of why you should try to acquire new evidence, was given by me in 1967, with acknowledgements to some earlier work. Nicod's criterion (pp. 53 and 164) was *succinctly* refuted by me (1967, 1968). On pages 60 and 86 there should have been a reference to 'evolving probabilities' (1967–8), perhaps better called *dynamic*, which change by thought alone instead of by empirical evidence, and are required for resolving important apparent paradoxes. Related to this matter are Axioms V* and VI* (p. 63) which are essentially my axiom A4′ (1947–50) and this is I think the most original part of my system of axioms (as distinct from their use). A detailed attack (p. 72) on Kolmogorov's axiom (countable additivity) was made by de Finetti (1949). The convergent sequence of probabilities 2^{-n} (p. 73) was proposed by Jeffreys for hypotheses containing n parameters ($n = 1, 2, 3, \ldots$). The notion of putting non-zero credibility at $p = 0$ and $p = 1$, where p is a physical probability (p. 78), is also due to Jeffreys, was adopted by me (1947–50) with some modification, and is one of the most original parts of Jeffreys's work (and is generalised in a natural manner to multinomials by Hintikka's system which is mentioned on p. 77). The use of a hierarchy of probabilities (p. 148) seems to be essentially the same as a proposal of mine made in 1951–2, was used very effectively for a statistical problem in 1965–7, and is now gaining prominence in Bayesian statistical literature.

On page 3, an event of zero probability is said to be impossible, and this conflicts with page 72. ('Almost impossible' would be better.) On page 2, $P(h|e)$ is expressed in terms of how probable e 'renders' h. But surely 'the moon is there' does not *render* $2+2 = 4$ probable. The proof on pages 73–4 that all precise physical probabilities not equal to 0 or 1 have zero credibilities is incorrect: a countable number of them could consistently be assumed to have positive credibilities. (The error is on p. 74, line 15.) Schlesinger's argument, treated at some length (pp. 84–5), contains a simple fallacy ($h_s \vee h_0$ is taken for granted in one part of the argument but not in another). Also Hempel's 'special consequence condition' (p. 87) is rather obviously incorrect (even the great make mistakes). On page 107 a criterion is given, with an acknowledgement to Blackburn, that enables one to regard 'green' as epistemologically superior to 'grue', namely, that to determine as well as one can that an object is grue necessitates doing the same for green. But this does not explain why the hypothesis of greenness is *far* superior. This is surely partly because the initial probability of grue suffers a great setback through the assigning of a *precise* date for the change from green to blue, apart from the arbitrariness of a sudden change.

Nelson Goodman's notion of 'entrenchment' in a language is mentioned in the chapter on simplicity, but Swinburne would rightly like some transcultural criteria. It seems to me that simplicity has to be taken *relative to some field of activity*, and then one would imagine an ideal language that would be most *economical* for that field. If the field is transcultural, then the ideal language might be also. The complexity of a proposition could then be defined as the length of the shortest expression of that proposition in this language. (My previous attempts to define complexity have been unsatisfactory and this one might be also.) Complexity, as thus defined, has some relationship with the initial credibility of the proposition, but the relationship remains obscure. For example '0 = 1' and '1 = 1' are equally simple but not equally probable. The proposition that 'there are chromosomes in all elephants aged 43 years' is more complex than 'there are chromosomes in all mammals', but has slightly higher initial credibility. We cannot decide which of these propositions is epistemologically superior by the Blackburn–Swinburne criterion. But in my judgment, the second proposition is of much greater explicativity for known observations because what little it loses on initial probability it gains back many times in what it explains.

Chapters 8 and 9 deal with the probabilities of specific propositions. In common with most philosophers of science, Swinburne does not mention in this context that he is talking about the estimation of probabilities in multinomial distributions, contingency tables, and multidimensional contingency tables, and he ignores the statistical literature on these topics. The only people I know, who have written on these subjects as if both statisticians and philosophers of science exist are Jeffreys (1939) and myself (1953-6, 1965, 1967). Neither statisticians nor philosophers have a prerogative to wisdom.

Although his philosophy is essentially Bayesian, Swinburne down-grades the concept of utility, and so never mentions, in the final chapter on acceptability, that the best theory for a specified purpose is in principle the one of maximum expected utility. Since utility is a subjectivistic concept, the best theory will depend on who you are as well as on the class of observations that are of interest to you at any given time. Moreover, in pure science utilities are difficult to estimate so that, to aid your judgment, it may be necessary to use quasi-utilities such as amount of information (in Fisher's or in Shannon's sense) or weight of evidence in Peirce's sense, or explicativity or predictivity in Good's sense. 'Explicativity', or the Sharpened Razor, synthesises some 'Bayesian' and Popperian ideas.

Although much thought has gone into the writing of this book it ignores too much of the literature for my liking.

I. J. GOOD
*Virginia Polytechnic Institute
and State University*

REFERENCES

GOOD, I. J. [1975*a*]: 'The Bayesian influence, or how to sweep subjectivism under the carpet' in W. L. Harper and C. A. Hooker (*eds*): *Foundations and Philosophy of Statistical Theories in the Physical Sciences.* Vol. 2: *Foundations and Philosophy of Statistical Inference.* Dordrecht: D. Reidel. (pp. 119–68).

GOOD, I. J. [1975*b*]: 'Explicativity, corroboration, and the relative odds of hypotheses', *Synthese*, **30**, pp. 39–73.

GILLIES, D. A. [1973]: *An Objective Theory of Probability*, Methuen & Co. Ltd. London. £3.50 Pp. x+250.

Gillies's aim is to present a new theory of probability that combines the mathematical attractions of Kolmogorov's measure-theoretic approach with the concern for empirical application evidenced by von Mises. The result—though new—is strangely archaic. Let me start with a small point which indicates Gillies's close reliance on frequentism. He rejects the Misean identification of probability with limiting frequency, on the grounds that it is too operationalistic, and tries to use a falsifying rule instead to establish a link between probability and frequency. Following Popper, from whom this quotation comes, he claims that such a rule is needed anyway because 'probability statements . . . [are] in principle *impervious* to *strict falsification*'. The usual argument, that in a coin-tossing experiment any arrangement of heads and tails has 'a finite probability . . . [and so] will not be strictly ruled out', is sketched on pages 161 f. But where is it derived that an event with finite probability is therefore a *possible* event? This is not a consequence of Gillies's theory, where probability is an undefined idea, neither a frequency or measure of propensity. Thus the argument fails to establish that a special rule is needed for falsification. To be sure, *something* has to be added to the theory. But it may be that we can add a further axiom that simply allows probability statements to be refuted in the usual way. (Such a theory is that outlined by Martin-Löf, *Information and Control* **9**, 1966, p. 616.)

The formulation of a falsifying rule is one of the principal topics of the book. More fundamental, however, are Gillies's treatment of independence, and his attempt, following von Mises and the early Popper (but not Reichenbach) in the frequentist tradition, to limit probability theory to sequences of independent events. Gillies, like Kolmogorov and most writers since, takes probabilities to be ascribed to events that occur under *repeatable conditions*. Using $m(A)$ for the number of times event A happens in n trials, and $p(A)$ for the probability, he states that a probability theory in which we did not have $m(A)/n \doteqdot p(A)$ at least with high probability 'would be extraordinary and most unacceptable' (see the first new paragraph on p. 80). He then proposes a set of repeatable conditions (Red-or-Blue) that are obviously not independent, and shows that in this case there is a rather low probability that $m(\text{Red})/n$ is anywhere near $p(\text{Red})$, even for large n. This leads to the 'axiom of independent repetitions', which Gillies suggests as a formal counterpart to von Mises's axiom of randomness. But it seems to me that neither of his arguments for this axiom is sound, and indeed that the axiom, like other attempts to contain the meaning or application of the word 'probability', is unnecessary and even undesirable.

The main argument (pp. 102f.), which concerns Red-or-Blue, goes like this. After each toss of a normal coin the event Red occurs if Tails are in the lead; Blue otherwise. Clearly, the occurrence of Red may be very heavily dependent on the result of the previous trials; the 4-tuple ⟨Red, Blue, Blue, Red⟩ for example is impossible (which, as we saw above, does not mean that it must have probability zero). Gillies, following Popper, feels that if the event Red has any sort of probability at all that probability must be $\frac{1}{2}$, in view of the symmetry of the game. He then quotes a result from Feller's book to the effect that in 30 million trials there is a very substantial probability that the relative frequency of Red will be nowhere near $\frac{1}{2}$—there is, for example, a probability of more than 0·6 that it will lie outside the interval $(\frac{1}{4}, \frac{3}{4})$. This is regarded as an 'intuitive contradiction', and the axiom of independent repetitions is brought in to rule it out. Though somewhat involved its intent is clear: probabilities relative to repeatable conditions are admissible only if the repetitions under these conditions are independent. But I would submit that this is not the actual force of the axiom at all.

For adding a new axiom cannot possibly *invalidate* a previously *valid* inference. If an 'intuitive contradiction' is validly derived without the axiom—in which case $p(\text{Red}) = \frac{1}{2}$ must be a premise—then it is validly derived with it. Thus either the inference is not invalidated (so that Axiom II is not sufficient), or it was already invalid (so that Axiom II is not necessary). Of course, what Gillies intends is that Axiom II should rule out not *derivations* but *interpretations of Axiom I* (the Kolmogorov axioms)—interpretations like Red-or-Blue. But this it fails to do.

It seems necessary to state the axiom in a little more detail. Gillies starts by telling us that 'probability theory deals with *probability systems* . . . $\langle \mathfrak{G}_s, \Omega, \mathcal{F}, p \rangle$ where $\mathfrak{G}_s$ is a set of repeatable conditions' and $\langle \Omega, \mathcal{F}, p \rangle$ is a probability space of the usual sort. (Here and elsewhere I have replaced Gillies's parentheses by the more distinctive French brackets.) $\mathfrak{G}_s^n$ is 'the set of conditions which are satisfied by a sequence of n repetitions of $\mathfrak{G}_s$' which are consecutive and non-overlapping (pp. 96 f.); Ω^n is the nth Cartesian power of Ω, $\mathcal{F}^n$ is a suitably defined σ-field of subsets of Ω^n; and p_n is the product measure. Axiom II now says (p. 103): 'If $\langle \mathfrak{G}_s, \Omega, \mathcal{F}, p \rangle$ is a probability system, so also are the quadruples

$$\langle \mathfrak{G}_s^n, \Omega^n, \mathcal{F}^n, p_n \rangle \text{ for all } n > 1\text{'}.$$

As an example, $\mathfrak{G}_s$ might describe the conditions of Red-or-Blue, and we might put $\Omega = \{\text{Red, Blue}\}$. With p as the uniform distribution we would have what is *prima facie* a probability system. Now set $n = 4$. Axiom II tells us that the structure $\langle \mathfrak{G}_s^4, \Omega^4, \mathcal{F}^4, p_4 \rangle$ is also a probability system. The probability of the impossible quadruple ⟨Red, Blue, Blue, Red⟩ is therefore equal to $1/2^4$. However, there is nothing contradictory about this conclusion, as I have already noted. Thus Axiom II does not in this way rule out Red-or-Blue from being a probability system. (Those who wish to demand—if necessary, by a further axiom—that impossible events have zero probability must replace Red-or-Blue by the very similar game Yellow-or-Black. We play this game by harnessing to Red-or-Blue a roulette wheel with a very large number of divisions—say 10^{1000}. If Red occurs, we spin the wheel; if 0 comes up then Black has occurred, otherwise Yellow has occurred. The wheel has the opposite effect in the case of Blue. In this game there are no impossible n-tuples, so Yellow-or-Black is not disqualified by Axiom II; nevertheless, an 'intuitive contradiction' is about as likely as it is with Red-or-Blue.)

Nor is there any other way that Axiom II can rule out Red-or-Blue. The axiom, after all, is satisfied if *every structure that satisfies Axiom I is counted as a probability system.* The axiom in fact is not a set-theoretical axiom for probability systems at all, but rather an axion for the 'universe of probability systems' (just as the ZF axioms are not axioms for sets, but for universes of sets). To this Gillies might respond that, contrary to what is suggested on page 104, a probability system is *not* a 'pure mathematical concept', but something concrete; that the axioms define not *possible* systems, but *actual* ones. But this still does not prevent all suitable structures from being probability systems, unless one takes account also of the falsifying rule; and if one does that, then the axiom is redundant. For the falsifying rule is very likely going to instruct us to reject the hypothesis $p(\text{Red}) = \frac{1}{2}$ which led to the 'intuitive contradiction' in the first place. In fact, we can put it like this: if the hypothesis is not rejected in a particular game, then the 'intuitive contradiction' does not arise.

To sum up: the axiom is either not sufficient or not necessary. There is no reason, of course, why Gillies should not adopt it; equally, he has given no reason why he should.

All in all, I think we can do very well without Gillies's relativization of probabilities to independent conditions. Indeed, I suggest that we should not even relativize probabilities to *repeatable* conditions. No law of probability theory, no law of large numbers, refers (or need refer) to repeatable conditions; most hold for sequences of mutually independent random variables (in the simplest cases only are they supposed to have a common distribution). The emphasis on repeatability (and to a lesser extent on independence) is natural if one is interested in *frequencies* (and therefore in testing), but pointless otherwise. In fact, if (as in some versions of the propensity interpretation) we take probability distributions as properties of individual situations, that is, *relative to the whole complex of conditions making up the situation* ('the whole state of the universe'), we can avoid all the difficulties with which Gillies struggles. By relativizing probabilities to a unique (therefore not repeatable) set of conditions we answer at once Gillies's objection (p. 153) that if we 'introduce probability for particular events . . . intuitively the probability will vary according to the set of conditions which the event is considered as instantiating'. Moreover, we show the irrelevance of his second argument (pp. 150–3) for Axiom II. This argument, in effect, says that we cannot introduce singular probabilities for events unless we know what repeatable conditions they are supposed to be relative to; and unless these conditions are independent we can easily obtain a variety of answers. Red, for example, in Red-or-Blue has intuitively a singular probability of 0, $\frac{1}{2}$, or 1, depending on what has happened before. But if, as I suggest, we do not require probabilities to be relative to repeatable conditions, we can easily allow $p(\text{Red})$ to vary from trial to trial.

Of the remainder of the book, which contains significant criticisms of, amongst other things, de Finetti's theory of probability, Neyman/Pearson testing, and the operationalist approach to 'conceptual innovation in the exact sciences' I do not wish to say very much. With much I agree. But there is one point which should be mentioned. On pages 193 f. Gillies discusses 'a paradox which apparently results from an adoption of . . . [the falsifying] rule'. This is that if the rule is applied once, and a refutation achieved, it may be possible so to reformulate the evidence, and reuse the rule, as to avoid a refutation. His solution, that 'observations' can be differently interpreted, seems beside the point, since there is no (synthetic) reinterpretation involved, merely a mathematical (analytic) transformation. Yet on pages 223–5, when discussing some arguments of Neyman, Gillies has a rather different attitude to what seems a very similar situation. (It is one of the few unclear parts of a book that, though stylistically a bit wooden, is remarkable for its clarity.) Neyman suggests that a statistician not working (as he recommends) with alternative hypotheses can find himself with evidence which reduced to one z-statistic refutes a hypothesis and reduced to another fails to do so. Gillies's reply is that two tests can be so related even in classical physics, so that Neyman's result cannot possibly show the need for alternative hypotheses (for clearly they are not needed in classical physics). Yet though these two problems can be made identical by a simple reformulation of each of them, Gillies nowhere connects the two passages.

The following minor errors and misprints could conceivably lead to confusion: page 15—'coherence' should read 'strict coherence'; page 19—'$r+1/n+2$' should read '$(r+1)/(n+2)$'; page 58—'$\geqslant$' on the 16th last line should read '$\leqslant$'; page 59—the law in question should read 'For rods or columns of a large number of different materials $\theta \propto l$, where θ is the temperature of the rod or column and l its increase in length'; page 98, line 16—'by the initial number n' should read 'with respect to the initial number n'; page 196, 12th last line—'a set k of parameters' should read 'a set of $k+1$ parameters'; pages 217 and 220—'$1/n(x_1+ \ldots +x_n)$' should read '$(x_1+ \ldots +x_n)/n$' and the expression for s^2 should be similarly rearranged; page 223—'x-test' should read 'z-test'; page 238—'the square of the difference between μ and a_1 $(\mu-a_1)^2$' would be more transparent as 'the square of the difference between μ and a_1, that is $(\mu-a_1)^2$'.

DAVID MILLER
University of Warwick

BUB, Jeffrey [1974]: *The Interpretation of Quantum Mechanics*. Dordrecht: Reidel. Dfl. 50. Pp. ix+155.

This book is an exposition of the current position of a major contributor to contemporary discussion on the foundations of quantum theory. Parts of it have appeared in a briefer form in some of Bub's recent articles but this is Bub's fullest and most self-contained account to date.

The first half of the book builds up to a perspicuous discussion of Kochen and Specker's results with a careful discussion of the inadequacies of the earlier von Neumann and Jauch and Piron no-hidden-variable proofs presented en route. The one chapter here that I am unhappy about is that containing

Bub's criticisms of the Bell–Wigner locality argument. Freedman and Wigner have in fact replied to an earlier but similar critique by Bub, and their reply (*Foundations of Physics* 3 (1973), 457–8) equally affects this chapter in Bub's book. However I understand that a further paper by Bub presenting a modified version of his critique will appear in a future issue of *Foundations of Physics*.

Bub's own putative solution of the fundamental problem in the interpretation of quantum theory—the measurement problem—is presented in the latter chapters of the book. I confess that these chapters left me somewhat baffled as to what Bub's position is.

Bub believes that quantum theory involves a non-classical logic in much the same sense as Kochen and Specker believe this. However Kochen and Specker only explicitly define *validity* for this non-classical logic, whereas in order to make its semantics unambiguous one needs definitions of *satisfaction* and *truth* as well as of validity. At one point (bottom of p. 126) it appears as though Bub is opting for an interpretation of this logic as a many-valued logic, but he has assured me that this is not in fact his position.

The most acute difficulty here is this: Bub wants measurements merely to discover and not to create the values of the measured observables. He therefore wants any measurable observable to have a particular (generally unknown) value at any given time whether or not the system is in an eigenstate of the corresponding operator. In particular each component of the spin of a spin-one system is to have a particular value at a given time. Now this is just what Kochen and Specker's argument seemed to rule out for a certain special choice of 117 different spin-axis directions. But Bub's position is apparently that it doesn't rule this out, but only rules out the conjunction of this with the assumption (rejected by Kochen and Specker) that the elementary propositions assigning values to spin-components obey the laws of classical logic.

Now it is not made clear in Bub's discussion just where his assumption of a non-classical logic for the propositions in question blocks the derivation of Kochen and Specker's contradiction, from the premises that he, Bub, accepts. He tacitly assumes that his interpretation of Kochen and Specker's result is the same as Kochen and Specker's interpretation. This seems to me unlikely because the assumption Bub needs to solve the measurement problem (a problem that Kochen and Specker never discuss) is equivalent to assuming that whenever $p \vee q$ is predicted by the quantum theory (and therefore, *a fortiori*, true) then either p is true (though not necessarily predicted) or q is true (though not necessarily predicted). Now certainly there is nothing in Kochen and Specker that commits them to this assumption (and some indication that they definitely reject it). And once you make this assumption it seems to me that the only way to avoid a contradiction, in Kochen and Specker's spin-one example, is to drop the inference from p is true and q is true to p and q is true. There are three difficulties about attributing some such proposal as this to Bub (*i*) one cannot decide from Kochen and Specker's incompletely prescribed semantics which inferences are and which inferences are not valid in quantum logic (no implication relation is defined), (*ii*) Bub would certainly wish the structure of his quantum logic to be imposed by specific features of the quantum theory and not subject to apparently free decisions on the part of the quantum logician, (*iii*) I've been quite unable to persuade Bub himself to accept this interpretation of his position.

I cannot thus understand Bub's putative solution of the measurement problem. However it is only fair to admit that Bub is not alone in defending his view (his position is shared by Bill Demopoulos), and Hilary Putnam's present position may be fairly similar (Putnam's original exposition of his view was seriously marred by his failure to take explicitly into account the implications of Kochen and Specker's results). Bub has an obvious debt to Putnam, which he has admitted handsomely in his previous writings, but Putnam is curiously not mentioned anywhere in the present book, apparently because 'the short bibliography lists only works directly cited, . . . since the sources of the ideas discussed will be obvious throughout'. But while Bub has allies, so have I. I could mention another worker in this field (well-known to readers of this journal) who even after extensive conversations with Bub and Demopoulos is just as baffled by their putative solution of the measurement problem as I am. I think therefore that we must await further clarification of Bub's quantum logic and of his putative solution of the measurement problem, before we can know whether his position is consistent, or indeed know more precisely what his position is.

JON DORLING
Chelsea College,
University of London

Cohen, Robert S. and Wartofsky, Marx W. (*eds.*) [1974]: *Logical and Epistemological Studies in Contemporary Physics* (Boston Studies in the Philosophy of Science, 13). Dordrecht: Reidel. Dfl. 55. Pp. viii+462.

Contents

Hüseyin Yilmaz: 'Perception and Philosophy of Science'
Jeffrey Bub and William Demopoulos: 'The Intrepretation of Quantum Mechanics'
C. A. Hooker: 'Defence of a Non-Conventionalist Interpretation of Classical Mechanics'
Laszlo Tisza: 'Comments on Hooker'
Bas C. van Fraassen: 'The Formal Representation of Physical Quantities'
E. J. Post: 'Comments on van Fraassen'
John Stachel: 'Comments on van Fraassen'
Bas C. van Fraassen: 'The Labyrinth of Quantum Logics'
Edward Mackinnon: 'Ontic Commitments of Quantum Mechanics'
John Stachel: 'Comments on Mackinnon'
Patrick Heelan: 'Quantum Logic and Classical Logic: Their Respective Roles'
A. R. Marlow: 'Implications of a New Axiom Set for Quantum Logic'
Milič Čapek: 'Two Types of Continuity'
A. Papapetrou: 'General Relativity—Some Puzzling Questions'
H. Bergman: 'Personal Remembrance of Albert Einstein'
H. Bergman (1929): 'The Controversy Concerning the Law of Causality in Contemporary Physics'

I cannot recommend this volume even for the college library.

Yilmaz's 90 page paper is described by the editors as a 'monographic essay in evolutionary epistemology, applying the conceptual and mathematical understanding of special relativistic quantum field theory to set forth a theory of the evolution and adaptation of perceptual structures'. It consists largely of summaries of some of the author's highly controversial contributions to mathematical psychophysics and to advanced mathematical physics.

Bub and Demopoulos is a respectable exposition of the authors' quantum logic position but by the time this volume appeared Bub had published an equally accessible, sharper, and much more complete exposition in his later book bearing the same title.

Hooker spends nearly seventy pages defending a thesis which seems to me so weak as to be scarcely controversial, namely that a realist interpretation of Newtonian forces can be consistently adopted but is not obligatory. Tisza's slight comments are not very germane and largely concern themselves with publicising a ten year old paper of his own on the logical structure of physics.

Van Fraassen's 'The Formal Representation of Physical Quantities' seems to be expounding the rather trivial thesis that some of the features of quantum logic would be preserved in a formal framework general enough to embrace other theories. E. J. Post's comments, on the doubtful validity of the point particle concept, are not especially germane to van Fraassen's paper. (He is not the only commentator in the volume who seems to have agreed to comment on the strength of a title which in the event proved misleading.) Stachel's comments, on the other hand, are relevant and incisive, and contain an important assessment of the virtues and defects of quantum logic and related approaches. Stachel also addresses himself to van Fraassen's second, much more valuable paper, 'The Labyrinth of Quantum Logics', which seems to have found its way into this volume as some kind of afterthought. This paper is an excellent and very clear analytical discussion of the various rival types of quantum logic and is a major pedagogical contribution to the literature. A library which obtained reprints of this paper and Stachel's comments would in my view have extracted everything of importance from the volume.

Mackinnon's discussions seemed to me alternately trivial and highly controversial, the latter parts being effectively dealt with in Stachel's comments.

Heelan's paper contains some rather facile criticism of Putnam's position together with the important claim that one can find logics with precisely the peculiar features of quantum logic in all sorts of fields in and outside science. Unfortunately Heelan's defence of this important claim is not sufficiently self-contained in the present paper for the reader to be able to assess its correctness. To do this he must turn to other papers of Heelan.

Marlow's paper is mainly concerned with announcing a remarkable mathematical result which purports to show (if I understand him aright) that for almost any reasonable classical theory there will be a quantum mechanical theory whose predictions include all the predictions of the classical theory. Unfortunately the proof of Marlow's putative theorem has merely been 'submitted for publication' in a reputable mathematical physics journal, and, so far as I have been able to ascertain, it has not to date appeared either in that journal or anywhere else. It seems to me unsatisfactory that philosophers should be expected to take cognisance of putative technical results which have not yet been accepted by a reputable specialist journal.

Capek's paper seems to be recommending that we replace the classical continuum in physics by something like the intuitionistic continuum, but this author is so mathematically unsophisticated that he even thinks that "continuity in the mathematical sense of the word is synonymous with *infinite divisibility*" (his italics).

Papapetrou's paper is a semi-popular exposition of matters which are well-known and contains no references.

The final paper by H. Bergman is merely an English translation of a work he published in 1929. It is largely a recital of other people's views at the time. Paul Forman's work has stimulated a certain historical interest in this area, but I cannot understand why Bergman's paper should be considered of any continuing philosophical interest, or why the editors should have heralded it as 'a minor classic in the philosophical understanding of contemporary physics'. It is prefaced by some rather naive personal reminiscences of Einstein.

It seems to me that, judging by this volume, we should ask our libraries to discontinue their automatic subscriptions to Boston Studies in the Philosophy of Science, and order, in future, only on the basis of inspection copies.

JON DORLING
Chelsea College,
University of London

HOOKER, C. A. (*ed.*) [1973]: *Contemporary Research in the Foundations and Philosophy of Quantum Theory.* Dordrecht: Reidel. Dfl. 56 paperback. Dfl. 105, hardback. Pp. xx+385.

Contents

Jeffrey Bub: 'On the completeness of quantum mechanics'
Leon Cohen: 'Joint probability distributions in quantum mechanics'
Bas C. van Fraassen: 'Semantic analysis of quantum logic'
E. Gerjuoy: 'Is the principle of superposition really necessary?'
R. J. Greechie and Stanley P. Gudder: 'Quantum logics'
C. A. Hooker: 'Metaphysics and modern physics: a prolegomenon to the understanding of quantum theory'
Arthur Komar: 'The general relativistic quantization programme'
F. David Peat: 'Quantum physics and general relativity; the search for a deeper theory'
Mendel Sachs: 'On the nature of light and the problem of matter'
Eugene P. Wigner: 'Epistemological perspective on quantum theory'

This book brings together a distinguished set of contributors discussing a single subject area. The collection will be an important reference source for workers in this field. That the papers contain few significant advances is to be blamed more on the difficulty of the central problems than on any lack of diligence on the part of the contributors.

Bub's paper contains a concise exposition of many of the theses of his more recently published book *The Interpretation of Quantum Mechanics.* He presents incisive criticisms of Heisenberg's 'measurement-disturbance' interpretation of quantum theory as also of the earlier 'no-hidden-variable' proofs of von Neumann and of Jauch and Piron. Bub's criticisms of Bell and the 'no-local-hidden-variables' literature are of a more controversial character. Bub himself opts for a quantum logic interpretation which seems to me not very different from orthodox Copenhagenism once we free the latter from mere measurement disturbances and from Bohr's Kantianism. Crucial to his interpretation is that we take Gleason's theorem and the Kochen and Specker 'no-hidden-variable' proof seriously, rejecting any attempts to get round these by the trick of supposing that there are vastly more observables in reality than are distinguished in the quantum theory. Bub's objection to the latter kind of hidden-variable theories (of which he was one of the original proponents) is that far from solving the measurement problem they simply shift it, whereas he holds that the measurement problem disappears once we take the quantum theory's own more economical assumption about the number of possible physical observables really seriously. He develops his own quantum logical 'solution' of the measurement problem in more detail in his book.

Cohen's paper expounds his own important negative results, on the possibility of hidden variables, but his article adds little to his earlier publications on the subject. It is nevertheless conveniently included in this collection.

Van Fraassen distinguishes, within a formal quantum logic and modal logic framework, two putatively consistent interpretations of the quantum theory. One of these is the kind of anti-Copenhagen hidden-variable theory which sidesteps the Kochen and Specker result by multiplication of observables, and which has already been discussed critically in Bub's paper. The other 'Copenhagen variant' is some distance from orthodox Copenhagenism and is what I should myself call a partial-hidden-variable theory. Van Fraassen does not suppose that all observables have definite values all the time, but nevertheless allows observables sometimes to have definite (but unknown) values when the system is not in a corresponding quantum mechanical eigenstate. In particular he supposes that when observables of two systems are correlated as in a measurement interaction, the correlated observables may be allowed to have definite but unknown values. This would enable us to say that mere looking at the pointer enables us to discover the pointer position rather than actually bringing into being a pointer position which did not exist before we looked. This interpretation would thus solve the measurement problem. However while van Fraassen correctly remarks that this view does not formally conflict with the theorem of Kochen and Specker (it does not constitute a hidden variable theory in the sense of those authors), it nevertheless seems to me that he overlooks a probable conflict with Gleason's theorem. It does not seem to me difficult to show that the latter creates the same difficulty for partial hidden variable theories such as van Fraassen's 'Copenhagen variant' as is created for fully-fledged hidden variable theories by the more restricted Kochen and Specker theorem.

Gerjuoy's paper is mainly a close assessment of the experimental evidence for the superposition principle, and particularly of comparatively recent experiments. His conclusion seems to be the unsurprising one that the evidence is pretty good.

Greechie and Gudder are mathematicians who are quantum logicians and the content of their paper is mainly technical. They do however make a number of controversial remarks about the hidden-variable problem *en passant*. They idiosyncratically pick on Jauch and Piron's argument as their paradigm of a no-hidden-variable theorem (Bub's paper, rightly in my view, rejected that argument as a non-starter) and they assure us that the investigators giving impossibility proofs are not referring to the same things as the hidden variable proponents. Their attempted demonstration of this is yet another presentation of the well-known type of hidden variable theory already considered in this volume by Bub and van Fraassen. It seems to me that historically they are quite wrong: that many hidden variable defenders would have hoped for just the kind of theory that Kochen and Specker rule out (and that previously von Neumann and Jauch and Piron had fallaciously purported to have already excluded): that the second-best hidden variable theories that they themselves characterise are notoriously in dire trouble with the 'no-local-hidden-variable-theory' results of Bell, Wigner, Clauser, Friedman, Horne, Shimony and Holt, although many of the latter authors would have welcomed such theories as rivals to quantum theory were they not ruled out by their own negative results. The suggestion that the provers of genuinely rigorous impossibility results and the physical speculators are merely at cross purposes seems to me quite the opposite of the true situation in this field.

The length of Hooker's paper is out of proportion to that of the other papers in the volume. He surveys various varieties of classical and non-classical theory from an abstract point of view which introduces a fair sprinkling of that forbidding, but increasingly standard, mathematical terminology which philosophers of science are now expected to take in their stride. His central suggestion is that we should look more in the direction of non-linear field theories than in that of traditional particle or field theories if we are to understand/replace the quantum theory.

Komar and Peat both concentrate on general relativity. In spite of their respective titles Peat's paper comes closest to a survey of philosophically relevant recent contributions and Komar's paper is a more personal report with recent references only to his own work. Peat favours a mildly holistic approach to the quest for a deeper theory. Komar thinks that his own preferred programme for quantising general relativity could well shed light on the reduction of the wave packet and on complementarity. I am hardly competent to assess these frankly speculative suggestions.

Mendel Sachs's paper is half semi-popular history, which is rather redundant in this collection, and half statements and claims about his own revolutionary non-linear theory which explains gravitation, electromagnetism, and quantum theory (if I understand him aright). Unfortunately the technical papers in which he has expounded his theory are forbidding, numerous, full of cross-references to each other, and none, unfortunately, published in journals whose standards of refereeing are entirely beyond reproach. A question mark therefore hangs over his work.

The final paper in the volume is by its most distinguished contributor and is such an excellent account of the problem situation that it could have been usefully placed at the beginning. Wigner formulates a number of problems about the interpretation of the theory, some less well-known ones as well as the usual ones, and states his views on each of the major proposals in the literature for solving the most central problems. He opts for an extreme form of instrumentalism as the least unacceptable consistent interpretation of the theory, but is obviously very unhappy about it.

JON DORLING
Chelsea College,
University of London

Recent Publications

BOOKS RECEIVED

ADAMS, E. W. [1975]: *The Logic of Conditionals.* Holland: D. Reidel Publishing Co. $22.00. Pp. 155.

ADORNO, T. W. *et al* [1976]: *The Positivist Dispute in German Sociology*, translated by G. Adey and D. Frisby. London: Heinemann. Hardback £6.50, Paperback £3.90. Pp. 307.

ANDERSON, A. R. and BELNAP, N. D. JNR. [1976]: *Entailment: The Logic of Relevance and Necessity.* New Jersey: Princeton University Press. £14.70. Pp. 542.

BENN, S. I. and MORTIMORE, G. W. (*eds.*) [1976]: *Rationality and the Social Sciences.* London: Routledge Kegan & Paul. £8.50. Pp. 416.

BLOCK, N. J. and DWORKIN, G. (*eds.*) [1976]: *The IQ Controversy.* New York: Pantheon Books. Pp. 557.

BLUMENBERG, H. [1975]: *Die Genesis der Kopernikanischen Welt.* Frankfurt: Suhrkamp Verlag. Pp. 803.

BOGDAN, R. J. (*ed.*) [1975]: *Local Induction.* Dordrecht: D. Reidel Publishing Co. $39.50. Pp. 338.

BREUSCH, F. L. [1976]: *Referat—Gegen Dogma und Dialektik.* Frankfurt am Main: Akademische Verlagsgesellschaft. Pp. 246.

CAPEK, M. (*ed.*) [1975]: *The Concepts of Space and Time.* Dordrecht: D. Reidel Publishing Co. Hardback $58.00, Paperback $19.50. Pp. 546.

CAPRA, F. [1975]: *The Tao of Physics.* London: Wildwood House Ltd. £6.95. Pp. 352.

CHADWICK, O. [1976]: *The Secularisation of the European Mind in the 19th Century.* Cambridge University Press. £6.50. Pp. 286.

COLBY, K. M. [1975]: *Artificial Paranoia—A Computer Simulation of Paranoid Processes.* Oxford: Pergamon Press. Pp. 113.

DOBBS, B. J. T. [1976]: *Foundations of Newton's Alchemy.* Cambridge University Press. £10.50. Pp. 300.

DONOVAN, A. L. [1976]: *Philosophical Chemistry in the Scottish Enlightenment.* Edinburgh University Press. £7.00. Pp. 343.

ENGELHARDT, T. JNR. and SPICKER, S. F. (*eds.*) [1976]: *Evaluation and Explanation in the Biomedical Sciences.* Dordrecht: D. Reidel Publishing Co. Pp. 240.

EVANS, G. and MCDOWELL, J. (eds.) [1976]: *Truth and Meaning.* Oxford: Clarendon Press. £11.50. Pp. 419.

GRENE, M. and MENDELSOHN, E. (*eds.*) [1976]: *Topics in the Philosophy of Biology.* Dordrecht: D. Reidel Publishing Co. Hardback $39.50, Paperback $19.50. Pp. 454.

HALLER, R. and GOTSCHEL, J. (*eds.*) [1975]: *Philosophie und Physik.* Wiesbaden: Friedr. Vieweg und Sohn. 24,80 DM. Pp. 138.

HANNAWAY, O. [1975]: *The Chemists and the Word.* Baltimore: The Johns Hopkins University Press. £6.60. P. 165.

Haraway, D. J. [1976]: *Crystals, Fabrics and Fields*. London: Yale University Press. £9.00. Pp. 231.

Harding, S. (*ed.*) [1975]: *Can Theories Be Refuted?* Dordrecht: D. Reidel Publishing Co. Hardback $39.50, Paperback $18.00. Pp. 366.

Harrison, J. [1976]: *Hume's Moral Epistemology*. London: Oxford University Press. £1.95. Pp. 131.

Hintikka, J. [1975]: *The Intentions of Intentionality and Other New Models for Modalities*. Dordrecht: D. Reidel Publishing Co. Hardback $45.00, Paperback $34.00. Pp. 262.

Hintikka, J. (*ed.*) [1975]: *Rudolf Carnap, Logical Empiricist*. Dordrecht: D. Reidel Publishing Co. $49.50. Pp. 520.

Hooker, C. A. (ed.) [1975]: *The Logico-Algebraic Approach to Quantum Mechanics—Vol. 1 Historical Evolution*. Dordrecht: D. Reidel Publishing Co. Hardback $68.00, Paperback $24.00. Pp. 607.

Icermann, L. and Penner, S. S. [1976]: *Energy: Vol II—Non-Nuclear Energy Technologies*. Massachusetts: Addison-Wesley Publishing Co. Inc. Hardback $19.50, Paperback $13.50. Pp. 673.

Johansson, I. [1975]: *A Critique of Karl Popper's Methodology*. Stockholm: Esselte Studium. Pp. 210.

Keenan, E. (*ed.*) [1975]: *Formal Semantics of Natural Languages*. London: Cambridge University Press. £12.50. Pp. 475.

Knorr, K. D., Strasser, H. and Zilian, H. G. (*eds.*) [1975]: *Determinants and Controls of Scientific Developments*. Dordrecht: D. Reidel Publishing Co. $58.00. Pp. 460.

Lackrill, J. L. (*ed.*) [1976]: *Aristotle's Posterior Analytics*. London: Clarendon Press. Hardback £7.50, Paperback £3.00. Pp. 277.

Lebesgue, H. [1975]: *La Mesure des Grandeurs*. Paris: Librairie Scientifique et Technique. 30 Fr. Pp. 184.

Lebesgue, H. [1975]: *Leçons sur les Séries Trigonométriques*. Paris: Librairie Scientifique et Technique. 28 Fr. Pp. 128.

Linstone, H. A. and Turoff, M. (*eds.*) [1976]: *The Delphi Method*. Massachusetts: Advanced Book Programme/Addison Wesley. Hardback $29.50, Paperback $16.50. Pp. 620.

Loeb, A. L. [1976]: *Space Structures: Their Harmony and Counterpoint*. Massachusettes: Addison-Wesley. Pp. 169.

Mach, E. [1976]: *Knowledge and Error*. Dordrecht: D. Reidel Publishing Co. Hardback $54.00, Paperback $32.00. Pp. 393.

Mackay, A. and Merrill, D. (*eds.*) [1976]: *Issues in the Philosophy of Language*. Yale University Press. £7.50. Pp. 161.

McClelland, P. [1975]: *Causal Explanation and Model Building in History, Economics and the New Economic History*. New York: Cornell University Press. £8.15. Pp. 290.

Mandler, G. [1975]: *Mind and Emotion*. London: John Wiley and Sons Hardback £7.65, Paperback £3.70. Pp. 280.

Manninen, J. and Tuomela, R. (*eds.*) [1975]: *Essays on Explanation and Understanding*. Dordrecht: D. Reidel Publishing Co. $56.00. Pp. 450.

Mehra, J. [1975]: *The Solvay Conference on Physics—Aspects of the Development of Physics since 1911*. Dordrecht: D. Reidel Publishing Co. $59.00. Pp. 415.

MITTLESTAEDT, P. [1976]: ***Philosophical Problems of Modern Physics.*** Dordrecht: D. Reidel Publishing Co. Hardback $13.50. Paperback $11.00. Pp. 211.

MURDOCH, J. E. and SYLLA, E. D. (*eds.*) [1975]: ***The Cultural Context of Medieval Learning.*** Dordrecht: D. Reidel Publishing Co. Hardback $49.00, Paperback $28,00. Pp. 566.

O'CONNOR, D. J. [1975]: ***The Correspondence Theory of Truth.*** London: Hutchinson University Press. Paperback £2.75. Pp. 144.

PARRINI, P. [1976]: ***Linguaggio e Teoria.*** Firense: La Nuova Italia. Pp. 314.

POLANYI, M. and PROSCH, H. [1976]: ***Meaning.*** London: MIT. £7.50. Pp. 246.

PUTNAM, H. [1975]: ***Mathematics, Matter and Method—Philosophical Papers Vol 1.*** Cambridge University Press. £7.75. Pp. 328.

PUTNAM, H. [1975]: ***Mind, Language and Reality—Philosophical Papers Vol 2*** Cambridge University Press. £9.25. Pp. 457.

RESCHER, N. [1975]: ***A Theory of Possibility.*** Oxford. Basil Blackwell: £8.50. Pp. 255.

ROBINSON, D. N. [1976]: ***An Intellectual History of Psychology.*** London: Macmillan Publishing Co. Pp. 434.

ROSEBERG, U. [1975]: ***Determinismus und Physik.*** Berlin: Akademie-Verlag, DM. 12.50. Pp. 208.

SCHMIDT, H. J. [1975]: ***Notes in Mathematical Physics.*** Marburg: Abteilung Mathematische Physik des Fachbereichs Physik der Universität Marburg. Pp. 131.

SHAFER, G. [1976]: ***A Mathematical Theory of Evidence.*** Princeton: Princeton University Press. Hardback £11.00, Paperback £5.60. Pp. 297.

SMART, B. [1976]: ***Sociology, Phenomenology and Marxian Analysis.*** London: Routledge & Kegan Paul. £6.00. Pp. 206.

SMITH, C. U. M. [1976]: ***The Problem of Life.*** London: Macmillan. £10 Pp. 343.

SOBER, E. [1975]: ***Simplicity.*** London: Clarendon Press. £6.50. Pp. 189.

SPINNER, H. [1976]: ***Pluralismus als Erkenntnismodell.*** Frankfurt am Main: Suhrkamp Verlag. Pp. 299.

SUPPES, P. [1976]: ***Logic and Probability in Quantum Mechanics.*** Dordrecht: D. Reidel Publishing Co. $39.50. Pp. 521.

THIBAUD, P. [1976]: ***La Logique de Charles Sanders Peirce.*** Aix en Provence: Universite de Provence. Pp. 184.

WEBER, R. L. [1975]: ***A Random Walk in Science.*** London: Institute of Physics. £4.75. Pp. 206.

WEIZENBAUM, J. [1976]: ***Computer Power and Human Reason.*** San Francisco: W. H. Freeman and Co. £5.95 ($9.95). Pp. 300.

WESTMAN, R. S. (*ed.*) [1976]: ***The Copernican Achievement.*** Los Angeles: University of California Press. £9.45. Pp. 405

WILLIAMS, C. J. F. [1976]: ***What is Truth?*** Cambridge University Press. £4.90. Pp. 102.

WISDOM, J. O. [1975]: ***Philosophy and Its Place in Our Culture.*** London: Gordon and Breach. $6.30. Pp. 270.

WOODFIELD, A. [1976]: ***Teleology.*** Cambridge University Press. £6.50. Pp. 232.

ZAKOPOULOS, A. N. [1975]: *Plato on* [illegible] ew York: Philosophical Library. $7.50. Pp. 142.

ZIMAN, J. [1976]: *The Force of Kn* [illegible] *Scientific Dimension of Society.* Cambridge University Press. [illegible] 3.

ZWART, P. J. [1975]: *About Ti* [illegible] North-Holland Publishing Co. Paperback Dfl. 50.00. Pp. 21[illegible]

Brit. J. Phil. Sci. 27 (1976), 307–315 *Printed in Great Britain*

501'02
B774

Scientific Laws that are neither Deterministic nor Probabilistic

by AIDAN SUDBURY

The purpose of this paper is to investigate the sorts of laws or hypotheses that systems may be explained by. We picture the scientist as performing a set of experiments (possibly infinite in number) and recording the results, and from these trying to determine the sort of mechanism that is involved. For example, he might be observing a system of particles and recording their positions (to a specified degree of accuracy) every second. Since the data received can be written down, it can be coded as a sequence of 0's and 1's. The scientific task, as idealised in this paper, is to find a hypothesis, finite in length, which explains this sequence. These hypotheses are to be framed in a particular language with only a finite number of symbols.

This paper will take a mathematical rather than a philosophical approach in that there will be a somewhat ruthless disregard of linguistic subtleties, and a good deal of legislation in order to make the ideas precise. The definition of 'explanation' given is essentially framed for probabilistic laws, and thus also for deterministic ones, since these are merely a special case. It will be shown that this definition allows a wider class of laws. These new laws will be theoretically testable in the same way as the more familiar ones, and thus are also candidates for laws of science, even if their interpretation is at the moment rather difficult.

It is relatively easy to say what it is for a sequence to be explained by a deterministic law. It is that there is an expression of finite length which for each $n = 1, 2, 3, \ldots$ specifies whether a 0 or a 1 occurs in the nth place. A sequence is explained by the law if it is the sequence specified. Put more mathematically, we have that a deterministic law is a computable function $f(\cdot)$ from the positive integers into $\{0, 1\}$ which explains the sequence $f(1)\,f(2)\,f(3)\,f(4) \ldots$. For example, we could have $f(n) = 1$ if n is prime, $f(n) = 0$ otherwise. The sequence explained by this function would then be 11101010001 . . . . It is obvious that to every finite sequence (to be called a string) of 0's and 1's there is a finite deterministic law that specifies it, simply the string itself. We shall thus only be concerned with explaining infinite sequences. We call the sequences explained by deterministic laws computable sequences. Since there are a countable number of laws and an uncountable number of infinite sequences and each deterministic rule specifies just one sequence, there must be sequences that are not computable.

Now it might be supposed that random sequences cannot be explained, but this is a misconception. Certainly they do not obey deterministic laws, but since there are tests for randomness, they do possess structure of a sort. It does make sense to say that a sequence has the characteristics of one derived from tossing a fair coin, when, for example, one thing to be expected is that the proportion of heads will tend to 1/2. The problem of what it is for a sequence to be explained by a statistical law is, however, not surprisingly rather difficult, and we shall mention three major attempts to solve it, although there have been many other contributions.

The first significant work done was that of von Mises. His idea essentially is that if one is gambling in a fair game then there should be no strategy that could give an expected gain different from 0. For example, in a series of independent trials with a fair coin where one loses one unit if it comes down heads and gains one unit if it comes down tails, whatever strategy one employs one's expected gain must be 0. His definition works well enough for independent trials, but is inadequate for the more complicated probabilistic laws we shall be considering.

Martin-Löf[1] has produced what is perhaps the most satisfactory definition of randomness (or obeying a probabilistic law), but it is both difficult to express in non-mathematical language, and less easy to extend than that of Howard,[2] with which we shall chiefly be concerned. Both place 'being explained by laws' firmly in the scientific context by defining the concept in terms of rejection rules, so let us begin by looking at a simple statistical test. Suppose we are testing a sequence of tosses of a coin for fairness and independence, and that our actual result is that the proportion of heads tends to 1/3. After 10 tosses the number of heads is 4, after 100, 30 and after 1000, 350. When, if ever, do we reject the hypothesis that the coin is fair? The answer is, of course, that it depends on how careful we wish to be. The probability of the number of heads in 10 tossings of a fair coin being as far from 5 (the expected number) as 4 is quite high, but it is extremely unlikely that it should be as far from 500 as 350. What we have to do is to decide what probability of making a mistake we are prepared to risk. Whatever probability we decide, however small, if we continue the trials indefinitely and the proportion of heads tends to 1/3 we shall eventually reject the hypothesis that the coin is fair. In this sense with an infinite sequence of outcomes we get an absolute rejection. There are, of course, many other tests for deciding whether a coin is fair or not. It is clear that the sequence of outcomes head, tail, head, ...

---

[1] Martin-Löf [1966].

[2] Howard [1975].

although possessing the right proportion of heads does not possess the right proportion of 'heads, heads', which should be 1/4 if the tosses were independent.

Martin-Löf defines a test in the following manner: first we set a level of significance, say 2^{-m}, then we prescribe for each n the 2^{n-m} strings of length n which are to be rejected by the test. Call the set of such strings U_m, then a test consists of a nested set U_i, $i = 1,2, \dots$ such that $U_m \supseteq U_{m+1}$ for $m = 1,2, \dots$ (that is, as the significance level increases, fewer strings are rejected). An infinite sequence s is rejected by U if the set of U_m that contain some initial segment of s has unbounded m, in other words if the sequence is rejected at an arbitrarily high significance level. We then consider the set of all tests that can be finitely defined (or are recursively enumerable) and say that a sequence is random if it survives all tests. Martin-Löf shows that almost all sequences are random (have measure 1 in the language of measure theory). What is remarkable is that under the definition of test given it appears impossible to devise a test of non-randomness. The tests mentioned here are suitable for testing for independence and fairness but they can be adapted for all probabilistic laws.

Martin-Löf claims that his definition is based on the work of Kolmogorov in defining randomness for finite strings. As has been pointed out, any finite string can be generated by a finite rule, but we still wish to say that certain finite outcomes are characteristically random whereas others are not. We wish to say that a table of random numbers is indeed random whereas a string consisting of 0's is not. Kolmogorov defines the 'complexity' of a string as the length of the computer programme required to generate it; for a string of 0's it would be short, for a table of random numbers it would be relatively long. The length of the programme is dependent on the computer used, but the longer the strings the less significant is this factor. Kolmogorov says that a string is 'random' if the programme required is of maximal length for strings of that length. It is found that nearly all strings are 'random' in this sense.

This seems to be the most satisfactory definition yet given for strings, but it is not clear that it does in fact capture the concept of 'randomness', except in so far as the term is restricted to cases in which 0 and 1 occur with equal probability. Firstly, consider the rule 0 'appears with probability 1/1000 and 1 with probability 999/1000'. This could generate a random string, but since there would be relatively so few 0's, it would require a shorter programme than a string characteristic of 0 appearing with probability 1/2. Secondly in his paper, 'Could there exist a world which obeyed no scientific laws?' Sudbury[1] gave a necessary criterion for a rule

---

[1] Sudbury [1973].

to be scientific, and with this restriction showed that if only a countable number of scientific laws are allowed (these include probabilistic laws) there will be infinite sequences of 0's and 1's which do not obey any of them. These sequences are extremely remote for even less can be determined of them than of the 'random' sequences. Possibly finite segments of many of them would be 'complex', and thus it would seem that the Kolmogorov definition will exclude many random sequences and include ones that are not.

No characterisation of what it was for a law to be 'scientific' was given by Sudbury, and it would be difficult or perhaps impossible to do so with such a vague term, but the question of whether there are any 'scientific' laws that are not deterministic or probabilistic is naturally raised and is the subject of the rest of the paper. We shall use Howard's definition of what it is for an infinite sequence to be explained by a probabilistic law or measure as our starting point.

A probabilistic law or measure is a function which says what the probability of each possible outcome is, that is, ascribes to each string of 0's and 1's a number. For example, if we had a series of independent trials with 0 appearing with probability $1/3$ and 1 with probability $2/3$ the sequence 001101 would have probability $(1/3)^3(2/3)^3$. If the probability of 0 in the first place was $1/2$ and subsequently the probability of a 0 being followed by a 1 was $2/3$ and that of a 1 being followed by 0 also $2/3$, then 001101 would have probability $1/2 . 1/3 . 2/3 . 1/3 . 2/3 . 2/3$. In fact any assignment M of numbers to the strings with the conditions that for any string s, $M(0)+M(1) = 1$, $M(s0)+M(s1) = M(s)$, and $0 \leqslant M(s) \leqslant 1$ where $s0$ is the string consisting of s followed by 0, is a probability measure. A computable measure is such an assignment M where M is a computable function. It should be noted that deterministic laws are special cases of computable measures, for given a computable function f, the associated computable measure gives the value 1 to the string $f(1)\,f(2)\ldots f(n)$ and 0 to all others of length n.

The Howard definition of 'explanation' is roughly this: a computable probabilistic law or measure explains an infinite sequence if there is no better explanation, 'better' being defined in terms of a likelihood ratio test. Consider two strings of length n, one that appears random and the other consisting entirely of 0's, why might we say that the first is explained by coin tossing whereas the second is not, when under that hypothesis both strings have the same probability 2^{-n}? The answer is, of course, that the hypothesis 'the sequence is just 0's' is better in that it gives the latter sequence a probability of 1.

More precisely the Howard definition (slightly modified) goes like this:

If there are two computable measures M and N and an infinite sequence s, then N is at least as good an explanation as M of s (written $M \leqslant {}_sN$) if and only if

$$\frac{M(s_1 \ldots s_n)}{N(s_1 \ldots s_n)} \text{ is bounded above for all } n,$$

where $s_1 \ldots s_n$ are the first n elements of the sequence s. (This differs from Howard who demands that the ratio tends to a limit).

Definition 1.

M explains s if and only if it is a greatest element in the partial ordering defined by $\leqslant s$.

Howard only allows computable measures as alternatives, and this is reasonable since they are the probabilistic laws that can be written down (trivially, to each sequence there exists a measure (not necessarily computable) which gives it the value 1). With this condition he shows that every computable measure M explains some sequences (in fact in the language of measure theory a set of M-measure 1). It is not difficult to show that any sequence which is random in a certain measure in the sense of Martin-Lõf is also explained by that measure in Howard's sense, but I do not know if the reverse is true.

Now, given any string of 0's and 1's a computable measure assigns a number (possibly 1) to it, the probability of that particular outcome if the string was being produced according to the rule expressed by the measure. These numbers are used to construct a test as to which explanation is most appropriate. We may wonder if this feature of assigning to each string a number which can be used to assess how well it is explained does not contain the essence of what a scientific theory should do.

We could try very tentatively the following definition: a scientific explanation of a sequence of 0's and 1's is a non-negative computable function with arguments the strings of 0's and 1's. Scientific explanations can then be tested using comparisons in the manner of the definition of $\leqslant_s$.

Now some sort of normalisation is needed for it is clear that, using tests like Howard's, a function $g(s) = |s| f(s)$, where $|s|$ is the length of the string s, will be better than $f(s)$ for all strings. An obvious normalisation is that for each n the sum of the numbers assigned to all strings of length n should be 1, but it might not in fact be the most appropriate.

To get new sorts of explanations we must break the additivity property of measures. If we specify that $M(s0)+M(s1) = M(s)$ for every string s then M is a measure. The new rules we seek must not have this property.

A class of 'scientific explanations' is now introduced and they will be called 'weights'. We say that a string s_1 of length $m+n$ extends a string s_2 of length n if s_1 has as its first n elements s_2, and we say that Ω^*, a set of strings $s_1, \ldots, s_n$, is a disjoint covering of Ω, the set of all strings, if every string in Ω is an extension of a string in Ω^* and no string in Ω^* is the extension of a different string in Ω^*. Let us call the members of Ω^*, $s_1, \ldots, s_n$ (they must be finite in number), then a weight W is a non-negative function defined on the strings such that there exist $\underline{w}$, $\overline{w}$ such that for all disjoint coverings Ω^*

$$0 < \underline{w} \leqslant \sum_{s_i \epsilon \Omega^*} W(s_i) \leqslant \overline{w} \leqslant \infty .$$

That is, the sum of the weights of a disjoint covering may not be arbitrarily large or arbitrarily small (for a probability measure $\underline{w} = \overline{w} = 1$). We show in the appendix that using tests like Howard's, restricting comparisons to computable weights, that every computable weight does explain some infinite sequences. However, the crucial question as to whether there are any sequences explained by the 'weights' that are not explained by probability measures has not as yet been answered. Nevertheless even without this, interesting theoretical questions are raised.

Firstly, is there any reason to suppose the non-additive weights introduced in this paper inadequate as scientific explanations since they can be tested with 'likelihood ratios' in the same way as statistical hypotheses would be rejected. Secondly, if they are adequate, how are they to be interpreted, for there is something strongly counter-intuitive about the way in which the weights are assigned. Since, given that a string s has appeared, the only possible next numbers are 0 or 1, it is hard to see how a rational system of weighting would not give $W(s) = W(s0)+W(s1)$.

But this difficulty really makes the understanding of such laws more interesting, for, although it is no way claimed as being such a radical step, it is worth remembering how great a conceptual jump there is from considering deterministic processes to considering those operating by chance. Any new kind of rule brings with it new ideas of how processes might operate.

This apparent irrationality may increase the difficulty of interpretation, but it also leads on to the only class of weights found which is not purely arbitrary. $W(s) \neq W(s0)+W(s1)$ has something in common with systems of logic in which $(A \wedge B) \vee (A \wedge \sim B) \not\equiv A$, and thus we look at a well-known type of many-valued logic. In this the truth-function T for the propositions has values in $[0,1]$ and $T(p)+T(\sim p) = 1$, $T(p \wedge q) = min(T(p),T(q))$. Now computable measures essentially assign numbers to conjunctions of statements, for when we put $M(010) = .05$ we mean that the probability that the first digit should be 0, the second 1 and the third

0 is .05. The many-valued logic given above gives an alternative method of assigning numbers to conjunctions. For strings it goes like this:

For each n we assign 2 positive numbers summing to 1 to the outcomes 0 or 1 in the nth place, and then calculate the unnormed weights of the strings taking the minima as in the rule given by the many-valued logic. For example:

	1	2	3	4	5
0		1/2	1/3	2/5	1/5
1		1/2	2/3	3/5	4/5

In this case $W(0) = 1/2$, $W(10) = 1/3$, $W(011) = 1/2$, $W(1101) = 2/5$. For each n we then sum the unnormed weights for the strings of length n and divide each by the sum.

It is not obvious that this method of assigning weights will satisfy the criterion in the definition of a weight, and in general it will not be easy to show that they do. For instance, if we let the numbers assigned to 0 be $1/2, 1/2^2, \ldots 1/2^n \ldots$, it is possible to construct a cover of Ω with arbitrarily small combined weight. The following example, however, is an assignment that does satisfy the criterion, and in a non-trivial way since the ratio of the weights of some strings to the sum of their extensions is unbounded above and not 1 as it would be for a computable measure.

Example

	1	2	3	...	n
0	1/2	1/3	1/4		$1/n+1$
1	1/2	2/3	3/4		$n/n+1$

This will not be demonstrated as the proof is complicated, and in any case this system of weighting is unfortunately equivalent to the measure appropriate to coin-tossing and therefore explains no new sequences. It should be remarked, however, that this system of 'weighting' is fairly simple, and so it is not clear that in general we could dismiss the new types of laws on the grounds of complexity.

If it could be demonstrated that there are sequences that are explained by the new laws but not by the old, then it means that it would be logically possible for a system to behave in a manner not explicable in terms of deterministic or probabilistic laws but explicable in terms of the extended class presented here. In that case the extended class of laws could not be ignored, and their interpretation would be of great interest, for the world might indeed be like that. More subtly, if it could be shown that there was a non-additive weighting such that the set of sequences explained by it could not also be explained by a finite number of probability measures, then it would mean that some of those sequences could only be explained by measures of arbitrarily great length, and thus the non-additive weight could be preferred on these grounds. If, on the other hand, it could be shown that the set of explicable sequences was not extended, it would be a step towards showing that the only laws we needed were the ones we have already, and this would be a surprising result.

It is hoped that this paper has raised several questions of interest and provided solutions to some of them. Perhaps the most compelling are still unanswered.

APPENDIX

Define Ω as the set of all infinite sequences of 0's and 1's, and ϕ the empty set. Let S^n be the set of all strings of length n. If s is a string then S_s is the cylinder set of all infinite sequences that have s as initial segment. For $|s| < n$, S_s^n represents the set of the strings of length n with s as initial segment (the extensions of s). A weight W is a function defined on the finite strings of 0's and 1's such that $W(s)$ is defined for each string s, and such that if a set of cylinder sets $S_{s_1}, \ldots, S_{s_n}$ has the properties that $\bigcup_{i=1}^{n} S_{s_i} = \Omega$ and $S_{s_i} \cap S_{s_j} = \phi$ for $i \neq j$, then for each W there exist strictly positive numbers $\underline{w}$ and $\bar{w}$ such that $\underline{w} < \sum_{i=1}^{n} W(s_i) < \bar{w}$.

[Note: the above properties imply the covering set is finite in number and for computable measures $\underline{w} = \bar{w} = 1$].

The partial ordering $\leqslant_e$ is defined for weights in the same manner as for computable measures.

LEMMA. *The sum of the weights of any set of disjoint cylinder sets is* $\leqslant \bar{w}$.

Proof. If it was greater than $\bar{w}$ then such a set would contain a finite subset with sum of weights $>\bar{w}$. This can be extended to a covering of Ω with sum of weights also $>\bar{w}$, thus contradicting the definition of a weight.

THEOREM. *Each computable weight W explains some sequences. The set of sequences not explained by W can be contained in a set of cylinder sets with arbitrarily small W-weight.*

Proof. Consider the computable weights $W_1, \ldots W_r, \ldots$ and let X_{W_i} be the set of infinite sequences s such that $\dfrac{W_i(s_1 \ldots s_n)}{W(s_1 \ldots s_n)}$ is not bounded above as $n \to \infty$. For each $s \in X_{w_i}$ and number K_i there thus exists a cylinder set S_{s^n} (where s^n is the string consisting of the first n elements of s) such that $\frac{W_i(S_{s^n})}{W(S_{s^n})} > K_i$. The set B_i consisting of all such S_{s^n} contains X_{w_i}, and further, since cylinder sets are either disjoint or one contains the other, B_i contains a subset B_i^* such that the union of the members of B_i is the same as that of B_i^* and the members of B_i^* are mutually disjoint. $\cup B_i^* \supset X_{w_i}$ and has combined W-weight $< \bar{w}_i/K_i$.

Thus X_w, the set explained by W is contained in a set of cylinder sets with combined W-weights $< \sum_1^\infty \bar{w}_i/K_i$ where the K_i may be chosen arbitrarily large. So X_W can be contained in the union of a set of cylinder sets with arbitrarily small combined W-weight, certainly less than $\underline{w}$ and thus $X_w \neq \Omega$.

Monash University

REFERENCES

Martin-Löf, P. [1966]: 'The definition of random sequences', *Information and Control*, 9, pp. 602–19.

Howard, J. V. [1975]: *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*. (21, 215–224).

Sudbury, A. W. [1973]: 'Could there exist a world which obeyed no scientific laws?', *British Journal for the Philosophy of Science*, 24 pp. 39–40.

Brit. J. Phil. Sci. 27 (1976), 317–328 *Printed in Great Britain*

An Integrated Biological Approach to the Species Problem[*]

by EROL F. GIRAY

Introduction.
1 *Desiderata for a Species Concept.*
2 *Morphological Species Concepts.*
3 *Biological Approach to the Species Problem.*
4 *An Integrated Biospecies Concept.*
5 *Some Objections to the Biospecies Concept.*
6 *Conclusion.*

INTRODUCTION

There has been much dispute in the recent literature regarding whether species are a 'real' and 'natural' unit of classification. Criticisms by taxonomists who claim that species are not real have had considerable impact upon taxonomy in the last two decades. The problem of defining a concept of 'species' which satisfies the desiderata for a natural unit of classification has been referred to as the 'species problem'. The purpose of this paper is (1) to show that a morphological definition of 'species' does not solve the species problem, (2) to introduce an integration of biological species concepts which have been proposed by biologists and philosophers, and (3) to prove that such an integrated approach fulfils the desiderata for a natural species concept.

1 DESIDERATA FOR A SPECIES CONCEPT

Species should be defined such that the rules to be applied in delineating species groups are coherent. We should never be led to classify two organisms both as belonging to the same species and as belonging to separate species. The rule (or rules) for classification must be clearly

Received 28 *September* 1975

---

[*] I wish to extend my appreciation to Kenneth S. Friedman and Warren M. Altkin for their encouragement and advice throughout previous drafts of the paper. I am also grateful for the helpful suggestions of the referees of this *Journal.*

stated. To illustrate this criterion, we may consider an example of an incoherent unit of classification, the 'order' of the Linnaean system of taxonomy. Some biologists refer to the Ungulata (hooved mammals) as an 'order', and classify Artiodactyla (even-toed ungulates) and Perissodactyla (odd-toed ungulates) as 'suborders'. Others consider the less inclusive groups as 'orders' and construe Ungulata as a 'superorder'. Neither approach is more correct or more natural than the other, for there is no rule of the form: 'Two organisms belong to the same order if and only if . . .'. Such a rule is needed for the species, however, because if no rule can be found, we might seriously doubt that the category has any significance as a scientific concept.

One might relax this requirement, so as to allow a species classification based on *two* rules, one specifying the conditions necessary for assigning two organisms to the same species, the other naming a set of sufficient conditions. However, these must be related, nomically if not logically, in such a way as to insure that no pair of organisms could satisfy the necessary conditions but fail the sufficient conditions, or satisfy the sufficient conditions while failing the necessary conditions.

It should also be possible to state the rule whereby a species classification is established in at least two logically independent alternative ways, which are interconnected by laws. This has been referred to by Schlesinger ([1963]) as 'Maxwell's criterion for scientific reality'. If the several alternative definitions of a concept can be interrelated by a network of laws, this would indicate that the concept represents some aspect of scientific reality. The ability of the biological species concept to satisfy Maxwell's criterion will be shown to be one of its strongest advantages.

Once a rule for species classification is established and applied, it should be possible to infer inductively some general characteristics of any particular species group on the basis of a sample of its members. Pratt ([1972]) emphasises this desideratum as the most important aim of a classificatory system. He writes:

> For science to be possible at all, individuals must be referrable to categories, and the categories must be such that a scientist must be able to think that when certain circumstances obtain in establishing facts about particular individuals he is thereby establishing facts about members of the category in general (Pratt, [1972], p. 318).

Hull ([1970]) points out, however, that we must not allow ourselves to fall into the trap of using characteristics to *define* the species, because by doing so, we would only be defining the names of individual species and not the general concept underlying species divisions. What is desired is that after the general rule for species delineation has been applied, the individual groups so delineated can each be described by a cluster (see Putnam, [1962]) of properties.

A species concept should be of sufficient generality to be applied uniformly to all organisms in which such groupings exist (see Lehman [1967]). If different rules of classification are needed for different groups of organisms, that would suggest that the rules were formulated *ad hoc* to deal with specific types of organisms. The species concept should be defined without regard to the particular organisms to which it is applied.

Finally, the extent of a species should not be *a priori* limited in space or in time. Lehman [1967] calls this the desideratum of spatial and temporal irrelevance. Most species are limited in space and time as a matter of fact, but this is not a nomically necessary consequence of a suitable species rule.

To recapitulate the desiderata for a species concept, them, it should be (1) expressed by a coherent rule, (2) definable in at least two ways which are logically independent but nomically convergent [*i.e.*, interconnected by laws], (3) amenable to inductive inferences, (4) uniformly applicable to all organisms, and (5) spatially and temporally uniform.

2 MORPHOLOGICAL SPECIES CONCEPTS[1]

By far the oldest concept of species is that of the *typological species*. Species were thought of as possessing discrete, unchanging archetypes comprising a number of essential morphological characteristics which identify it and distinguish it from all other species. Every species was thought to continue in its original form indefinitely, any variation within the species being incidental. This definition dissolves the problem of temporal uniformity, as it entails species constancy across time. As such, however, it conflicts irresolubly with evolutionary theory, in which species are anything but constant. Moreover, if characteristics can in fact be classified as 'essential' or 'incidental' (and this is highly questionable), then there is no reason that the same characteristics should remain essential when the natural selection pressures on a species change. Few modern scientists would argue for retention of a typological species concept.

The *morphological* species (morphospecies) concept results from the typological concept once the notion that certain characteristics are 'essential' is purged (Lehman [1967]). Two organisms are said to be conspecific (*i.e.*, of the same species) if and only if their phenotypes (overall morphological characteristics) are sufficiently similar. Organisms which share clusters (*sensu* Putnam) of covarying characters are classified together; the smallest such groups are species. Ruse ([1969], [1973]) defines the morphospecies as:

---

[1] 'Morphology' is used here in a broad sense, so as to encompass anatomical characters as well as behaviour and other aspects of the phenotype.

> a concept based on the assumption that organisms can be grouped into taxa on the basis of some kind of *'overall' morphological similarity* and difference. Morphological species, according to the concept, are the smallest collections of organisms possessing this kind of overall morphological similarity, separated from other organisms by a *distinct morphological gap*, and containing no such internal morphological gap separating the members of the species from each other (Ruse, [1973], p. 127; my italics).

There is no coherent rule that can be used to determine whether two organisms belong to the same species, and thus the morphospecies concept fails the first desideratum. How similar do organisms have to be before they are similar 'overall'? How large is a 'distinct gap'? The morphospecies proponent cannot determine, for a given case, whether he is dealing with a single morphospecies (containing major overall similarity and indistinct gaps) or separate morphospecies (with lesser similarities and a distinct gap).

Suppose, for instance, that the morphospecies proponent is considering domestic dogs (species *Canis familiaris*), and that he has already separated this group from wolves, foxes, cats, rabbits, *etc.* The only rule provided him by the morphospecies concept is to continue dividing the group until no distinct gaps remain within the final groups. He must thus classify poodles, German shepherds, *etc.* in separate species. Certainly, clusters of characteristics may be found for most canine pedigrees. To multiply this absurdity further, in a great many instances the male and female sexes are distinct, each having a multitude of morphological properties which constitute a gap between the two sexes. Obviously, the smallest group which contains no distinct gaps in overall morphological characteristics will be far smaller than what one would wish to designate as a species.

It seems that when one decides that the classification of organisms has reached the species level, one must be tacitly applying some other rule than the incoherent one contained in the morphospecies definition. The biologist must rely upon a more tangible relationship than 'similarity' or 'difference' if he is to stop the species-splitting operation before the poodles are split from the shepherds and the females from the males (or even if he is to recognise the undesirability of basing species distinctions on those grounds). It is here that the biological species concept comes to use. Although the data from which one can draw inferences may be 100 per cent morphological, the critical question—are these organisms conspecific or not?—can be answered only by using a rule which relates to biological interactions between the organisms.

3 BIOLOGICAL APPROACHES TO THE SPECIES PROBLEM

It is not surprising that most modern biologists who subscribe to the viewpoint that species are natural units hold some variety of biological species concept (biospecies) as the solution to the species problem. Several such varieties have been expounded (see Simpson [1951], Mayr [1963] and [1969]; Lehman [1967]; and Pratt [1972]).[1] Although the interconnection of these various concepts has generally been overlooked (and will be demonstrated in the next section), some of the common postulates are ostensible: (1) the organisms of a biospecies are biologically interrelated with one another in a way that they are not related with organisms of other biospecies, (2) these interrelations transcend morphological similarity, and (3) biospecies, once separate, can virtually never be reunited. None of the biospecies proponents deny that information about the morphology of organisms is useful and often necessary as a clue to the intrrelations implied in a biospecies concept. Morphological data, however, are not the *definiens* of biospecies. It is rather the genetical and or evolutionary relations which are considered important.

Simpson, a paleontologist, prefers a biospecies definition which emphasises the evolution of species through time. According to Simpson, an 'evolutionary species' is an 'ancestral-descendant sequence of populations evolving separately from others and with its own unitary evolutionary role and tendencies'. His definition stresses timewise change, an aspect of the species which a neontologist such as Mayr would probably find of secondary importance. In his [1969], Mayr defines species as 'groups of interbreeding natural populations that are reproductively isolated from other such groups'. Although reproductive isolation is inferred from phenotypic evidence, the major thrust of the biospecies concept is on genotypic discontinuity between species and genotypic continuity within species—for it is genes which are transmitted from one organism to another.

Other biospecies proponents are even more explicit in their reference to the genotype as the basis of biological species. Lehman relates the concept of species with that of the gene pool, defined in terms of breeding populations. As applied to sexually reproducing organisms, a breeding population is one in which the probability of an individual mating with one member of the opposite sex is the same as with any other member. A gene pool consists of all breeding populations among which transfer of genetic material occurs under natural conditions. Both Mayr and Lehman construe the species as a group of interbreeding populations, and is difficult to see how their concepts differ, apart from the choice of terminology.

---

[1] These are not the only biospecies definitions which have been advanced in recent decades, but they have been among the most influential in biological thought.

Pratt discusses the species in terms of genetic interchangeability, *viz.*, a species consists of organisms whose genotypes are such that, in general, an organism will be viable even if one or more of its genes be replaced by the corresponding allele of any other organism of the species.

4 AN INTEGRATED BIOSPECIES CONCEPT

It can now be shown that the gene-pool concept of Mayr and Lehman, the gene-interchangeability concept of Pratt, and the evolutionary species concept of Simpson are logically independent but nomically convergent. As such, the integrated biospecies concept fulfils Maxwell's criterion (the second desideratum).

Let us first examine the connection between the gene-pool concept and the interchangeability concept. We may consider the genotype of any biparentally conceived organism as if it were a replica of one of the parental genotypes, with a random sample (generally amounting to one-half) of the genetic material replaced by genes from corresponding loci of the other parent's genotype. If the organisms of a species are to interbreed successfully, the viability of the recombinant genotypes must be sufficient to allow reproductive processes to sustain the species. Hence, interbreeding implies genetic interchangeability, and Pratt's definition follows. Moreover, biparental reproduction is highly adaptive, because of the greater flexibility of evolution as compared to strictly uniparental reproduction (see Mather [1973]). If a group of organisms is genotypically similar to the extent that their genes are frequently interchangeable, then evolutionary theory leads us to expect that they will interbreed, so as to spread the genes responsible for adaptive variants throughout the group. Hence, genotypic interchangeability implies interbreeding, and Pratt provides a logically independent alternative to the gene-pool concept, which can be linked to it by laws which follow from commonly accepted biological theory.

To relate Simpson's evolutionary concept to the gene-pool species, we note that if two populations do not share a common gene pool, one would expect that each would evolve separately and have its own evolutionary tendencies. Several aspects of evolutionary theory indicate this. For one, adaptive variants which arise by mutation in one species and spread through that species are not transfused to another; random mutation alone would lead to divergence in the absence of interbreeding. Second, because speciation (the division of one species into two or more) is often caused by geographic isolation,[1] incipient species are susceptible to different environmental selection factors, and so tend to evolve separately. Hence, genetic discontinuity implies evolutionary discontinuity, and Simpson's definition follows. Moreover, if two populations evolve separately, different mutations must be occurring in the two. But if the populations shared a gene pool, the effects of these mutations would be shared; thus, the populations must have separate gene pools. Hence, evolutionary discontinuity implies genetic discontinuity, and Simpson also provides a logically independent alternative to the gene-pool concept, which can be linked to it by laws which follow from commonly accepted biological theory.

At this juncture, it is worthwhile to reconsider the morphospecies concept in light of Ruse's ([1969], (1973]) contention that it too can satisfy Maxwell's criterion. Ruse's argument may be summarized as follows: The organisms of a biospecies share a fundamental genotypic similarity among themselves. This leads one to expect that the co-members of a biospecies will be morphologically similar, because the adaptive variants within the species will be spread throughout it, expressing themselves as common morphological adaptations. Hence, biospecies should be morphospecies. Moreover, it is advantageous for co-members of a morphospecies to interbreed (in order to disseminate to many the adaptive characteristics of a few), and also advantageous for members of different morphospecies *not* to interbreed (in order to avoid producing hybrids maladapted to both parental niches). Hence, morphospecies should be biospecies apparently linked to them by laws which follow from commonly accepted biological theory.

It is immediately apparent from Ruse's argument that the biospecies concept satisfies our third desideratum, *viz.*, permitting inductive generalisations from a few individuals to the species. Organisms that share a gene pool, should, in a rough sense, have similar phenotypes. This enables us to make certain inductive inferences about the species. There will be exceptions to any inferences we attempt to make, but these will be sufficiently few that the endeavour remains worthwhile. Knowing that an organism belongs to species *S* provides information which increases the probability that the organism possesses certain morphological characteristics, as compared to the prior probability that any randomly chosen organism is similarly endowed. As Ruse ([1969]) has pointed out, it is possible, given a biospecies, to draw several inductive generalisations, which constitute a cluster. Although not every member of the species *Homo sapiens* will necessarily be a big-brained featherless biped, the probability is much higher than it would be for an organism from any other biospecies.

---

[1] Mayr ([1963]) states that in nature geographic isolation is a *necessary* condition for speciation.

Though Ruse's argument supports the biospecies concept, it does not help to justify the morphospecies definition. Certainly, biological theory leads us to expect that genotypic similarity should be approximately reflected in phenotypic similarity, especially if the organisms interbreed. But whereas the genotypic similarity can be translated to a rule for biospecies classification, the phenotypic similarity cannot be translated to a coherent morphospecies classification rule.[1] In contrast to the biospecies, the morphospecies does not have an independent coherent definition.

5 SOME OBJECTIONS TO THE BIOSPECIES CONCEPT

One of the most obvious criticisms of the biospecies concept (at least as it is defined by Mayr and by Lehman) is that it cannot be applied to uniparental species, and thus fails to satisfy the desideratum of uniformity (the fourth desideratum). This objection is popular among morphospecies proponents and supporters of numerical classification methods, and Ruse raises it to bolster the claim that there is no advantage of the biospecies over the morphospecies. The problem, so the argument goes, is that uniparental organisms are reproductively isolated from everything else, and therefore, each must be classified as its own biospecies. Mayr concedes the difficulty, and suggests using the biospecies criterion for biparental organisms and defining the species in some other way for uniparental organisms. The critics reject this approach as *ad hoc*, and rightly so. However, the occurrence of strictly uniparental reproduction is far lower than was previously believed. Biparentalism may occur even in species whose primary mode of reproduction is uniparental. The only necessary condition for organisms to share a gene pool is that they can each contribute a part of a descendant organism's genotype by recombination. Many unicellular species effect this recombination by conjugation, in which two organisms come into mutual contact and exchange portions of their genotypes before separating. In many bacterial species, genetic material can be transmitted from one organism to another by phage viruses; this process is known as ***transduction.*** (See Stent [1972] for a detailed discussion of these and other mechanisms of genetic exchange.) It is clear that the gene pool is a useful scientific concept even for predominantly uniparental organisms. Since it is advantageous for genotypically similar organisms to interbreed, it is not difficult to see why almost all species have mechanisms whereby genetic continuity can be maintained.

---

[1] Ruse might appeal to Pratt's concept of genetic interchangeability, and answer with a criterion of 'phenetic interchangeability'. That is, he might define a morphospecies as a group of organisms any of which would be viable even if one or more of its phenotypic characters were replaced by the corresponding character of another member. But one cannot define 'one or more characters', because the phenotype cannot be divided into a finite number of non-overlapping 'phenes' in the way that the genotype can be divided into genes. Furthermore, it does not make sense to speak of interchanging phenetic characters, while interchange of genes has a very real context in the recombination of alleles which occurs in biparental reproduction. One cannot refer to a character of one organism 'corresponding' to that of another, except insofar as they are produced by corresponding alleles of the genotype. Thus, one is still forced to rely upon the biospecies concept for the interpretation of morphological information.

Another objection which has been raised against the biospecies concept, which in fact must be dealt with by any species concept, is the classification of *Rassenkreise* ('rings of races'). There are many instances in nature in which a series of populations P_1, P_2, . . ., P_n exists such that any P_i can interbreed with P_{i-1} and with P_{i+1}, but the terminal populations P_1 and P_n cannot interbreed although they overlap geographically. If each organism is to be classified unambiguously into one and only one species, there are three options: (1) to include the entire *Rassenkreis* in a single species even though the terminal populations do not interbreed, (2) to classify each population as a separate species even though the adjacent populations do interbreed, or (3) arbitrarily divide the group into k species, where k is more than 1 but less than n. It appears that all three options violate the desideratum of uniform applicability to all organisms. But if we consider the meaning of 'gene pool', it becomes clear that by assigning the entire *Rassenkreis* to a single species, this desideratum is satisfied. If P_1 and P_2 interbreed, this implies that genes from the former can be infused into the latter. The progeny of this intergression can then interbreed with members of P_3, and these progeny would in turn have genetic material derived from all three populations. Even though a combination of P_1 and P_3 genes alone might be inviable, the combination of all three contributions could very well be successful. This can be extended to more complex *Rassenkreise,* although the spread of an adaptive variant from one terminal population to the other is slower for larger rings. Hence all populations in a *Rassenkreis* are genotypically continuous, and should be classified in a single species.

A third problem for the biospecies concept (which again applies to any attempt at species classification) is that of the temporal continuity of evolution. With minor exceptions (*e.g.*, allopolyploidy), each organism is genotypically continuous with its immediate ancestors and descendants. Given the solution which has just been proposed to the *Rassenkreis* problem, we might consider every ancestral-descendant sequence of populations as a *Rassenkreis,* and furthermore, all descendants of a common ancestor would belong to the same species. It would seem that the discontinuity between species which is generally so ostensible when only contemporaneous organisms are considered vanishes when the time dimension is introduced. If we attempted to solve the problem by assigning the only contemporaneous organisms to the same species, we would violate the criterion of temporal uniformity, and would imply that whenever an organism dies, the species to which it belonged would need to be renamed.

Is the taxonomist confronted with a giant *Rassenkreis?* Fortunately not, because one of the premises in the argument which compels one to assign a contemporaneous *Rassenkreis* to a single species becomes false when the organisms are not contemporaneous. In the former situation, P_1 and P_n are conspecific because the intermediate populations make it possible for a P_1 organism and a P_n organism to have common descendants. If the intermediate populations are extinct ancestors, though, they cannot serve as the medium through which their descendants might have common progeny.

Supposing, then, that we have an infinite number of species classifications, each limited to one instant in time, we can arrive at a satisfactory scheme whereby temporally uniform species can be set up. The operating rule would be: Two temporally limited species, S_1 existing at time t_1, and S_2 existing at a later time t_2, are conspecific if and only if S_2 is the only species existing at t_2 whose members are all descendants of S_1, and no species descended from S_1 became extinct before t_2. According to this rule the common ancestor of two species is conspecific with neither. When one species splits into two, the original species, for all intents and purposes, is extinct.

This solution, arbitrary though it may appear, can be derived in a logically independent way based on Simpson's evolutionary species concept. Suppose we begin with a species S_1, from which evolves a single species S_2, and that from S_2 emerge two species S_{3a} and S_{3b}. Both S_1 and S_2 tend toward an evolutionary path which will lead to speciation producing S_{3a} and S_{3b}. Therefore, S_1 and S_2 are time strata of the same species. Neither S_{3a} nor S_{3b} is evolving in the direction of speciation to produce S_{3a} and S_{3b}; if either of these undergoes fragmentation of the gene pool, altogether different species result. S_1 and S_2 satisfy the criteria for assignment to a common species, but S_2 and S_{3a} (or S_2 and S_{3b}) do not. The same solution which was proposed in the preceding paragraph now is seen to follow from the biospecies concept itself (of which Simpson's 'evolutionary species' is only one facet), and is not at all an *ad hoc* answer to the problem.[1]

---

[1] Ruse ([1973]) claims that Simpson's evolutionary species is not spatially or temporally uniform, because in introducing evolution into the species concept, he is requiring the species to be restricted to descendants of organisms at a particular place and time. An

6 CONCLUSION

The biological species concept certainly fulfils the desiderata for a 'natural' or 'real' scientific concept. It can be clearly defined in a coherent rule for species classification. It has at least three logically independent definitions which are interconnected by biological laws. Inductive inferences from individuals to species are possible. There are few, if any, organisms to which the biospecies concept cannot be straightforwardly applied. Finally, it is in no way restricted spatially or temporally.

Armed with a scientifically sensible species concept, we might proceed to attempt to find natural ways of grouping these species into higher taxa. Pratt [1972] suggests an approach based on a conceptualisation of taxa as hierarchically arranged systems consisting of subsystems at lower levels of organisation. This theory has some appeal, though it is not clear that the levels of organisation of the systems are discrete or finite in number. Yet it seems that once we have grouped organism into natural non-overlapping species, the major problem of taxonomy has been solved, and it makes little difference for theoretical considerations how the species are grouped. One may be interested primarily in phenotypic similarity, or phylogenetic descent, or other types of relationships. Before any of these groupings makes sense, however, one must have natural species groups with which to work. An integrated biological species concept provides the most promising approach.

State University of New York at Oswego

REFERENCES

HULL, D. L. [1970]: 'Morphospecies and Biospecies: A Reply to Ruse', *British Journal for the Philosophy of Science*, **21**, pp. 280–2.

LEHMAN, H. [1967]: 'Are Biological Species Real?', *Philosophy of Science*, **34**, pp. 157–67.

MATHER, K. [1973]: *Genetical Structure of Populations*. London: Chapman & Hall.

MAYR, E. [1963]: *Animal Species and Evolution*. Cambridge, Mass.: Belknap.

MAYR, E. [1969]: *Principles of Systematic Zoology*. New York: McGraw-Hill.

PRATT, V. [1972]: 'Biological Classification', *British Journal for the Philosophy of Science*, **23**, pp. 305–27.

PUTNAM, H. [1962]: 'The Analytic and the Synthetic', in H. Feigl and G. Maxwell (*eds.*): *Minnesota Studies in the Philosophy of Science*. Minneapolis: University of Minnesota Press, pp. 358–97.

---

evolutionary species on Earth, so the argument goes, could not admit organisms from outer space, or from prehistory. But in fact, this is only *a posteriori* knowledge. If man colonises other planets but maintains genetic continuity with the race back at home, Simpson's definition would agree with others in that the colonisers were conspecific with the Earthlings. If Earth species are indeed limited to Earth, it is only because outer space seems to contribute an exceptionally strong isolating factor.

RUSE, M. [1969]: 'Definitions of Species in Biology', ***British Journal for the Philosophy of Science***, **20**, pp. 97–119.
RUSE, M. [1973]: *The Philosophy of Biology*. London: Hutchinson & Co.
SCHLESINGER, G. [1963]: *Method in the Physical Sciences*. New York: The Humanities Press.
SIMPSON, G. G. [1951]: 'The Species Concept', ***Evolution***, **5**, pp. 285–98.
STENT, G. [1972]: *Molecular Genetics*. San Francisco: W. H. Freeman.

Brit.J. Phil. Sci. 27 (1976), 329–362 *Printed in Great Britain*

Ad Hoc Auxiliary Hypotheses and Falsificationism

by ADOLF GRÜNBAUM*

1 *Introduction.*

2 *Three Historically Relevant Senses of* 'ad hoc *Auxiliary Hypothesis*' *in Ascending Order of Logical Strength.*

3 *Is There a Purely Logical, Non-pragmatic Characterisation of* ad hoc *Auxiliary Hypotheses?*
  (*a*) *Critique of Popper's Falsificationist Account of Legitimating* ad hoc *Auxiliaries.*
  (*b*) *Hempel's Thesis That 'There is, in fact, no precise criterion for* ad hoc *hypotheses*'.

4 *Is There a Clear Normative Rationale for Derogating* ad hoc *Hypotheses Methodologically as a Cognitive Liability?*

1 INTRODUCTION

In the history of science, there are various instances of theory-*modification* by *auxiliary* hypotheses, as distinct from *wholesale* theory-*replacement*. And these instances include familiar episodes in which the new auxiliaries prompted cries of *ad hoc*, epistemic *caveats* or at least a clear appreciation of considerable epistemic risk from some segments of the scientific community. In each such episode, a modifying collateral assumption H repaired a troubled major theory T_1 by then enabling the ensuing modified version T_2 to explain deductively a known fact E which is contrary to T_1. Thus T_2 is logically *incompatible* with T_1.[1] But the new auxiliary H was

---

* The author thanks Professors Carl G. Hempel, Laurens Laudan, Jarrett Leplin, Bernard R. Goldstein, and Victor Thoren for useful critical comments and/or references to historical literature and is indebted to the National Science Foundation (U.S.A.) for the support of research.

[1] Leplin ([1975], p. 317) unfortunately speaks of an original theory as being 'supplemented' by an auxiliary H when he lays down the following of his necessary conditions for ad hocness: 'If an hypothesis H is introduced into a theory T in response to an experimental result E, then if H is *ad hoc*, E is anomalous for T but not for T as supplemented by H'. This usage of 'supplemented' is unfortunate for the following reasons: (*i*) Leplin himself is rightly careful to note (*ibid.*, pp. 326–7) that whenever E's anomalousness derives from its incompatibility with T, the conjunction of T with H will be *inconsistent*, but (*ii*) as shown by Leplin's use of the terms 'supplementation' and 'supplementing' on his pp. 323 and 325 respectively, to 'supplement' a theory is to give a *consistent extension* of it, so that T as supplemented by H would have to be consistent after all for *every* anomalous E. Leplin's use of 'supplemented'—as contrasted with the use of

deemed cognitively a *deus ex machina* and hence methodologically somewhat suspect or even quite illegitimate. Cases in point which readily come to mind just from the physical sciences alone are: The modifying ascription of *negative* weight to phlogiston, the initial postulation of an extra-Uranian planet, Leverrier's hypothesis of one or more intra-Mercurial planets, the contraction hypothesis of Lorentz and Fitzgerald, and the assumption of the existence of neutrinos. Laudan ([1977], p. 117) mentions Ptolemaic astronomy, Cartesian physics and phrenology (Cantor [1975]) as having furnished other classic instances of *ad hoc* modifiers. In *some* such episodes, the postulation of H was justly held to deserve no censure but only a *caveat* as follows: Before H can be taken seriously as a cognitively viable alternative to some (tacit) contrary logical constituent of T_1, and thus before T_2 can be expected to remain a better theory than the refuted T_1, the procurement of 'additional' or '*independent*' empirical support for H is desirable or needed. But there were other episodes in which the espousal of H was deservedly treated much more pejoratively and was charged with outright epistemic illicitness and/or methodological turpitude.

I shall begin by trying to *adumbrate* three historically relevant senses of '*ad hoc*' as applied to *auxiliaries*, and by exhibiting the hierarchical logical connections among them. Thereafter, I shall consider the availability of a more precise articulation of these senses in purely logical—as distinct from pragmatic—terms. And I shall comment on some attempts to *justify* the cognitive disparagement and/or methodological interdiction of auxiliaries which are *ad hoc* in one or more of the three specified senses. Let me emphasise, however, that I do *not* claim the latter senses to be the only historically relevant ones. Moreover, note that by confining the discussion to *ad hoc* auxiliaries, I am concerning myself here with only one species of the genus of *ad hoc explanations*. For this genus also comprises theories which are held to be *ad hoc* individually *without* relation to a troubled earlier theory.[1]

'modified'—is thus infelicitous in one of his necessary conditions for *ad hocness*. But he pointedly demands consistency of H with only the 'essential' propositions of T when stating another of his necessary conditions (*ibid.*, p. 327).

Whereas Leplin was *not* victimised by his own infelicitous usage of 'supplemented', I have explained in my ([1976*a*], pp. 8–9) how Popper's misleading conjunctive description of the planetary τ_2 as 'consisting of the original theory plus the auxiliary' (Popper [1974], Book I, p. 33, item (*e*)) misled him to claim falsely that τ_2 has greater informative content *IC* than τ_1.

[1] Popper ([1972], p. 192) illustrates the different species of theories which are thus *non*-relationally *ad hoc* by means of the following purported explanation of the fact that the sea is very rough today: Neptune is very angry today. Popper deems the theory ascribing anger to Neptune today to be an *ad hoc* explanation, because the roughness of today's sea is presumably the *sole* evidence for it.

2 THREE HISTORICALLY-RELEVANT KINDS OF *ad hoc* AUXILIARIES

Auxiliary hypotheses H *alone*, in isolation from logically stronger theories, typically do not have any logical consequences which scientists would categorise as 'observational' in the pertinent context of inquiry. Even our putative empirical finding E is not a logical consequence of H alone in any given case but rather of the wider theory T_2. And hence, for every typical H, taken by itself H is also devoid of 'observational' entailments *other than* E, even if H is *not ad hoc* in any of the three senses below. Nonetheless, as an *ellipsis*, the statement that a given H does have testable consequences which differ both from E and from the implications of T_1 would be quite sound under the following conditions: The theory T_2—or some other theory which is likewise logically much stronger than H (besides asserting appropriate initial conditions)—does have test implications that depend on H but which do *not* belong to either the Tarskian consequence class of T_1 or to that of E. It is to be understood that in formulating various senses of '*ad hoc* auxiliary', I shall sometimes speak *elliptically* by talking about the empirical import of H rather than of T_2.

In contexts in which H might be charged with being *ad hoc* in one or more of our three senses below, the postulation of H is *prompted solely* by the aim of explaining the otherwise anomalous observed finding E amid preserving the main postulates of T_1. And therefore H is specifically *tailored* to enable the modified version T_2 of T_1 to accommodate E explanatorily, thereby essentially 'rescuing' the *core* of T_1. Hence it has been considered a cognitive or methodological desideratum or imperative that besides E, H have some '*additional*' actual or potential support not possessed by T_1. This imprecise desideratum has given rise to the likewise imprecise demand that, in the context of T_2, H have either tested or testable consequences which are intuitively '*independent*' in *at least* the following sense: They differ logically not only from E and E's entailments but also from the consequences of T_1, as well as from H itself. And, very roughly speaking, in the *weakest* of the three senses of '*ad hoc*', an auxiliary which fails to have independent consequences that are already tested would be considered as '*ad hoc*', provided that it likewise fails to have either the sanction or the disapprobation of other accepted theoretical principles.

For now, I shall merely *adumbrate* three historically relevant senses of '*ad hoc*'. And in doing so, I can therefore invoke the rough intuitive sense of the concept of independent consequence. But we shall see in Section 3(*b*) that the already stated attributes of an intuitively independent consequence fail to be defining characteristics of such a consequence, if there are to be any auxiliaries at all which are *ad hoc* on the strength of *not* having independent testable implications.

I shall refer to the three senses of '*ad hoc*' which I am about to discuss by appending respectively '(*a*)', '(*b*)' and '(*c*)' as a suffix to '*ad hoc*'. And we shall see that the corresponding attributes sustain unilateral entailment relations such that being *ad hoc* (*c*) is logically the strongest of the three while being *ad hoc* (*a*) is the weakest.

Auxiliaries Which Are ad hoc (*a*) *At a Time* t_0.

An example will serve to introduce my characterisation of this kind of auxiliary hypothesis.

When discussing the 'immunisation' of a theory from a falsification by the legitimate introduction of a modifying auxiliary, Popper ([1974], Book I, p. 32) cites the paradigm example of the postulation of the planet Neptune, which served to explain observations of the motion of Uranus that had been incompatible with the prior version of Newtonian planetary theory. And Popper comments on the methodological moral of this episode by saying:

> . . . we must not exclude all immunizations, not even all which introduce *ad hoc* auxiliary hypotheses. For example the observed motion of Uranus might have been regarded as a falsification of Newton's theory. Instead the auxiliary hypothesis of an outer planet was introduced *ad hoc*, thus immunizing the theory. This turned out to be fortunate; for the auxiliary hypothesis was a testable one, even if difficult to test, and it stood up to tests successfully (*ibid.*).

Note that Popper speaks of the Neptune hypothesis as having been 'introduced *ad hoc*'. But he points out that besides being independently *testable*, this auxiliary has *since* actually been tested, and indeed successfully. Moreover, the independent optical testability of the hypothesis was *known* at the time when Adams and Leverrier first propounded it: A suitably directed telescope was expected to detect Neptune. *In principle*, this planet could further make its presence felt by slight gravitational effects on planets *other than* Uranus, so that there is also *independent* gravitational testability in principle. Pertinent historical material is given in section 4 below.

In order to render the sense in which the auxiliary H of Adams and Leverrier can be said with Popper to have been 'introduced *ad hoc*', let me first explain some notation. Let τ_1 be the *pre*-Neptunian version of Newtonian planetary theory of *circa* 1820 which asserted that besides Uranus (and the Asteroids) there are *only* six planets all of which are *intra*-Uranian, while τ_2 is the amended Adams and Leverrier version of solar planetary theory which asserts the existence of the additional trans-Uranian planet Neptune. Then the auxiliary hypothesis H is the assertion that this additional planet exists *and* exerts non-negligible gravitational influences on the motion of Uranus. Furthermore, if U_1 is the 'basic' statement asserting τ_1's *incorrect* prediction of the orbit of Uranus, and U_2 is the *contrary* basic statement codifying the actually 'observed' orbit of Uranus, then τ_2 entails U_2, which thus plays the role of our paradigmatic E above, while τ_1 entails a statement which corresponds to the negation of E. Thus, τ_1 and τ_2 are just a special case of our incompatible ancestor and rescuing theories T_1 and T_2. Furthermore, here I shall retain the notation of my ([1976*a*], pp. 2 and 6–7), and hence '$LC(T)$' will refer to the Tarskian logical content or consequence class of the theory T. Furthermore, '$KC(T,t)$' will denote the presumably *proper* subset of $LC(T)$ whose member-statements are *known* to belong to $LC(T)$ at the time t by the pertinent segment of the scientific community.

Then if t_0 is roughly the time (month or year) when Neptune was first postulated, we can say that $KC(\tau_2, t_0)$ did *not* contain any independent testable consequences of τ_2 which had either *already* been *tested* by the time t_0 or which were then known to have the sanction or disapprobation of an already accepted other theory. Nor did the planetary H itself—which does *not* qualify as an 'independent' consequence of τ_2—have such sanction or disapprobation. Given this state of affairs, the concept of *ad hoc* (*a*) to be defined below will enable me to say that our planetary auxiliary H was '*ad hoc* (*a*)' at the time t_0 of its enunciation. And thus I suggest that the concept *ad hoc* (*a*) serves to explicate Popper's remark that H was 'introduced *ad hoc*'.

Speaking generally again, bear in mind hereafter that our intuitive concept of the set S of '*independent*' testable consequences of T_2 *excludes* from S H itself as well as the members of $LC(T_1)$ and the members of $LC(E)$. Also note that, for brevity, I shall speak of a given consequence C of a theory T_2 as 'already having *empirical sanction* at a time t', iff C has *either* already been successfully tested by the time t *or* is then known to have the sanction of an already accepted *other* theory. By the same token, if C has already been experimentally refuted by the time t or is then known to have the disapprobation of an already accepted *other* theory, I shall say that C already has '*empirical disapprobation*' at t. Finally, observe that T_2 is neither an accepted nor a refuted theory when the auxiliary H is first propounded.

Using the imprecise, *enthymematic* intuitive concept of an *independent* logical consequence of T_2, we can now give the following definition:

(*Def. a*): An auxiliary H of a 'rescuing' theory T_2 is *ad hoc* (*a*) at time t with respect to the corresponding earlier 'rescued' theory T_1 and a prior empirical finding E, iff at time t, H neither has the sanction, nor is rendered doubtful by any accepted theory *and* $KC(T_2,t)$ contains no testable *independent* consequences which already have either empirical sanction or disapprobation. Here the characterisation of T_1 and T_2 as respectively 'rescued' and 'rescuing' is intended to have the force of the following requirement imposed by Leplin ([1975], p. 327): H must be consistent with the retained *'essential'* propositions of T_1, where an 'essential' proposition of a theory T is one whose rejection 'would count, in the judgment of the scientific community, as a rejection of T irrespective of the retention of other hypotheses'. But I shall explain in Section 4 below why I do not impose Leplin's so-called 'condition of non-fundamentality' ([1975], p. 331) as one of the necessary conditions in the *definiens* of any of my three senses of '*ad hoc*'.

Here I now need to comment on the logical 'decomposition' of T_1 into a conjunction $T_0 \cdot M$ such that (*i*) each of the two conjuncts T_0 and M is logically *weaker* than T_1, (*ii*) T_0 is the '*core*' theory which is rescued by T_2 from the demise of T_1, but (*iii*) M is *repudiated* by T_2's assertion of a new auxiliary H which is logically contrary to M. Thus, if '$\rightarrow$' and '$\leftrightarrow$' respectively denote logical implication and logical equivalence, we have

$$T_1 \leftrightarrow T_0 \cdot M, \quad \text{and} \quad T_2 \leftrightarrow T_0 \cdot H,$$

where each of the two conjunctions are presumed to be internally consistent.

Professor Carl Hempel has suggested to me (in personal correspondence) that the logical 'part' T_0 of T_1 which is left after excision of M and which is then retained for incorporation in T_2 might be more specifically characterised by means of the following considerations. Letting the tilda '$\sim$' denote negation, we know that $T_1 \rightarrow (\sim M \vee T_1)$, $T_1 \rightarrow M$, and $[(\sim M \vee T_1) \cdot M] \rightarrow T_1$. Hence we can express T_1 as follows:

$$T_1 \leftrightarrow [(\sim M \vee T_1) \cdot M]$$

Hempel then points out that the *intersection* of the Tarskian consequence classes $LC(M)$ and $LC(\sim M \vee T_1)$ contains no factual statements but only tautologies. For the intersection is just the Tarskian consequence class of the *logically true disjunction* $M \vee (\sim M \vee T_1)$. In short, there is *complete* factual non-redundancy between the assertion $\sim M \vee T_1$ and the part M of T_1 which is to be excised by T_2. Given the latter kind of complete non-redundancy, Hempel suggests that the core part T_0 of T_1 which is to be retained by T_2 could be identified as equivalent to $\sim M \vee T_1$, so that

$$T_2 \leftrightarrow [(\sim M \vee T_1) \cdot H].$$

Nonetheless, Hempel wonders whether the specific construal of T_0 as $\sim M \vee T_1$ throws any light on actual historical cases. For he suspects that, in the latter cases, M has 'tentacles reaching to every corner of T_1', which I interpret to mean at least that the intersection $LC(M) \cap LC(T_0)$ is *not* factually empty after all.

Let me now use the concrete historical case of our two *planetary* theories τ_1 and τ_2—which are, of course special cases of T_1 and T_2 respectively—to give concrete substance to Hempel's suspicion and to comment on the parsimonious logical decomposition of T_1 into the conjuncts M and $\sim M \vee T_1$.

If we write $\tau_1 \leftrightarrow \tau_0 \cdot M$ and $\tau_2 \leftrightarrow \tau_0 \cdot H$, then I suggest that the following construals of M, H and τ_0 would be *roughly* appropriate to the given astronomical episode:

(*a*) M asserts at least that there are *no* trans-Uranian planets which would exert *non*-negligible gravitational forces on Uranus,

(*b*) H asserts at least that there is *exactly one such* trans-Uranian planet ('Neptune') of specified gravitational properties, and

(*c*) τ_0 asserts at least Newton's laws of motion and gravitation as well as that there are no less than six *intra*-Uranian planets. Given the latter construals, I observe the following:

1. The putative rendition of τ_0 as $\sim M \vee \tau_1$ would turn the disjunction $M \vee \tau_0$ into a *logical truth* and thereby would render the intersection $LC(M) \cap LC(\tau_0)$ factually empty. But in the actual historical construal which I have outlined, the disjunction $M \vee \tau_0$ far from being a logical truth is presumably an *empirical falsehood*. And thus, unlike $LC(M) \cap LC(\sim M \vee \tau_1)$, the intersection $LC(M) \cap LC(\tau_0)$ is not factually empty.

2. The rendition of τ_0 as $\sim M \vee \tau_1$—and, more generally, of the core T_0 by $\sim M \vee T_1$—allows $LC(T_0)$ to contain *only* (and all) those member-statements which belong to *both* $LC(\sim M)$ *and* $LC(T_1)$. In particular, this putative rendition *excludes every non-tautological consequence of* M from membership in $LC(T_0)$. But no such restriction seems to have been imposed in the historical case of the planetary τ_0.

3. In actual historical cases, there would seem to be no insistence on the kind of parsimonious logical decomposition of T_1 into conjuncts T_0 and M in which there is a *total lack* of factual redundancy between the conjuncts.

So much for my comments on the logical decomposition of T_1 into T_0 and M.

It is now clear that when first propounded, the Neptune hypothesis was *ad hoc* (*a*), as was Leverrier's postulation of the elusive planet Vulcan to account for the motion of Mercury. Moreover, we saw that although at that time t_0, $KC(\tau_2,t_0)$ thus did not contain any independent consequences which *already had* empirical sanction or disapprobation, $KC(\tau_2,t_0)$ did contain *other*, as yet untested though testable independent consequences. Yet we can well imagine a different auxiliary which is not only *ad hoc* (*a*) at the time t_0 of its genesis, but whose $KC(T_2,t_0)$ does not even contain any such other untested consequences, *i.e.*, whose $KC(T_2,t_0)$ contains no testable independent consequences *at all.* Hence we now consider a species of auxiliaries which are more strongly *ad hoc* than *ad hoc* (*a*).

Auxiliaries Which Are ad hoc (*b*) *At a Time* t_0

In the definition I am about to give, I assume that when an independent consequence of T_2 does belong to $KC(T_2,t)$ at time t, then C is known to be independent in addition to being known to belong to $LC(T_2)$.

(*Def. b*): An auxiliary H of a 'rescuing' theory T_2 is *ad hoc* (*b*) at time t with respect to the corresponding earlier 'rescued' theory T_1 and a prior empirical finding E, iff at time t, H neither has the sanction of nor is rendered doubtful by any accepted theory *and* $KC(T_2,t)$ contains no testable *independent* consequences.

Clearly, any auxiliary which is *ad hoc* (*b*) at time t will also be *ad hoc* (*a*) at that time. But as illustrated by the case of the Neptune hypothesis, in general the converse does not hold.

However, as shown by the debate between E. Zahar, P. K. Feyerabend, A. I. Miller and K. F. Schaffner in the pages of this *Journal* (**24**, pp. 95–123 and 223–62; **25**, pp. 25–78), it is at least arguable that *at some stage*, the Lorentz–Fitzgerald contraction hypothesis—hereafter '*LFC*'—was *ad hoc* (*b*) rather than merely *ad hoc* (*a*). And more confidently, one can say that the auxiliary postulation of the existence of neutrinos in the theory of radioactive nuclear disintegration was *ad hoc* (*b*) when it first took shape in the mind of Pauli *circa* 1930, well before the 1956 detection experiment of C. Cowan and F. Reines was even envisioned, and before the neutrino was known to have other theoretical sanction.

Human limitations being what they are, one may safely presume that for any sophisticated, high level scientific theory T_2 and at any stage t of actual science, even disregarding the already infinite set of T_2's tautological entailments, $KC(T_2,t)$ will be only a *proper* subset of $LC(T_2)$ rather than identical with it. Yet there are auxiliaries that have been the target of *ad hoc* charges which are stronger than *ad hoc* (*b*) allegations by being predicated on a daring claim about the membership of the entire class $LC(T_2)$ rather than merely about its proper subclass $KC(T_2,t)$. Hence we now consider the meaning of these charges under the rubric '*ad hoc* (*c*)'.

Auxiliaries Which Are ad hoc (*c*) *At a Time* t_0

(*Def. c*): An auxiliary H of a 'rescuing' theory T_2 is *ad hoc* (*c*) at time t with respect to the corresponding earlier 'rescued' theory T_1 and a prior empirical finding E, iff at time t, H neither has the sanction of nor is rendered doubtful by any accepted theory *and* $LC(T_2)$ contains no testable independent consequences.

Note that the time-dependence of the *ad hoc* (*c*) attribute is assured by the time-dependence of the *acceptance* of a given theory in the scientific community: The latter time-dependence is ingredient in the *first* conjunct of the *definiens* of *ad hoc* (*c*).

Let 'S' denote the set of T_2's *independent* testable consequences, while '$\subset$' refers to the *proper* subset relation. In virtue of the aforementioned limitations of human knowledge and the logical strength of typical sophisticated scientific theories T_2, we may assume that for most auxiliaries and at any time t, not only $KC(T_2,t) \subset LC(T_2)$ but also

$$S \cap KC(T_2,t) \subset S.$$

Of course, in the utopian event that the classes $KC(T_2,t)$ and $LC(T_2)$ are identical, the latter proper inclusion will not hold. Nor will it hold, for example, in the rare case of an auxiliary H for which even S is *empty* and not just $S \cap KC(T_2,t)$: Such a rare H is *ad hoc* (*c*) rather than only *ad hoc* (*b*). Since the emptiness of S obviously entails the emptiness of its intersection with $KC(T_2,t)$, any *ad hoc* (*c*) auxiliary must also be *ad hoc* (*b*). But clearly, an auxiliary which is *ad hoc* (*b*) at a given time will *not* necessarily also be *ad hoc* (*c*).

Precisely because at any given time, S is typically a very much wider class than $S \cap KC(T_2,t)$, a provably *ad hoc* (*b*) auxiliary can only be *conjecturally ad hoc* (*c*). By the same token, in the context of a sophisticated scientific theory T_2, it is incomparably harder—if feasible at all—even to *propound* an interesting *demonstrably ad hoc* (*c*) auxiliary than one which is demonstrably *ad hoc* (*b*).[1] Nonetheless, the very strong *ad hoc* (*c*) charge has been repeatedly levelled against the *LFC* auxiliary. Thus, at one time Popper ([1959*a*], p. 83) levelled it within a normative methodological framework in which *ad hoc* (*a*) auxiliaries need not necessarily be even peccadilloes—as we saw *à propos* of the Neptune hypothesis—but in which both *ad hoc* (*b*) and *ad hoc* (*c*) collateral hypotheses presumably commit a kind of mortal rather than venial sin. After stating his own special falsificationist requirements for methodologically admissible auxiliaries—about which more below in section 3—Popper writes:

---

[1] Agassi ([1975], ch. 8) rightly considers the nub of this conclusion to be one of the morals implicit in my earlier writings on *ad hoc* hypotheses (Grünbaum [1959], [1964] and [1973], ch. 12, section D, ch. 21, and Appendix, §§ 34–5, pp. 834–9). Agassi's chapter 'Towards a Theory of *Ad Hoc* Hypotheses' is addressed to a conception of an *ad hoc* auxiliary which he characterises as follows:

> Let us assume that what is rational is what has a general rationale rather than has no better cause than a whim. Now the *ad hoc* is neither due to a whim nor due to a general rationale. And so it can be either irrational, an excuse to stick to a rationale which turns out to be less general than originally assumed, or non-rational, either in being an exception or in being a *sui generis* or anything in between, e.g., a case of an unstudied territory (p. 189).

> An example of an unsatisfactory auxiliary hypothesis would be the contraction hypothesis of Fitzgerald and Lorentz which had no falsifiable consequences but merely served to restore the agreement between theory and experiment—mainly the findings of Michelson and Morley. An advance was here achieved only by the theory of relativity which predicted new consequences, new physical effects, and thereby opened up new possibilities for testing, and for falsifying, the theory ([1959*a*], p. 83).

In a critical note in this *Journal* about this passage of Popper's, I pointed out (Grünbaum [1959], pp. 48–50) that the *LFC* has 'a falsifiable bearing on the outcome of the Kennedy–Thorndike experiment' (p. 49), an experiment interestingly different from the Michelson–Morley type. And, as Zahar ([1973], pp. 104–5) has noted in his summary, elsewhere (Grünbaum [1964] and [1973], ch. 12, section D; ch. 21) I gave a further analysis whose pertinent upshot is the following: (*i*) In the case of the 1932 Kennedy–Thorndike experiment, the *LFC* predicts a specific *positive* shift of the interference fringes quantitatively *different* from the positive effect predicted by the original, unmodified aether theory and also from the *null* effect predicted by the special theory of relativity ('*STR*'). The prediction of a *positive* shift which the *LFC* makes concerning the outcome of the Kennedy–Thorndike experiment counts as an '*independent*' consequence of the Lorentz–Fitzgerald modification of the aether theory, because this positive prediction contrasts with the prediction of a null shift which the *LFC* makes for the Michelson–Morley experiment. (*ii*) The outcome of the Kennedy–Thorndike experiment had the capability of either confirming or falsifying the positive prediction concerning it made by the *LFC*, although the experimental falsification of this *prediction* does *not* suffice to falsify the *LFC itself*, since the *LFC* invoked other assumptions of the aether theory in making the prediction (Grünbaum [1964], p. 1411 and [1973], pp. 722–3). (*iii*) In virtue of the stated independent testability of the *LFC*, the latter has never been *ad hoc* (*c*).

In amplification of (*ii*) of these contentions, I cite the following statements from my 1964 article in *Science*:

> It is evident that it is logically possible for a Kennedy–Thorndike type of experiment to *confirm* the quantitative predictions of the Lorentz–Fitzgerald hypothesis

> as against those of the original aether theory independently of the Michelson–Morley experiment.
>
> Furthermore, it is a matter of *empirical* fact that the Kennedy–Thorndike experiment of 1932 did *not* yield a shift in the interference fringes corresponding to the time difference [variation] deduced from the Lorentz–Fitzgerald hypothesis. In fact, just like the Michelson–Morley experiment, the Kennedy–Thorndike experiment had a negative outcome in the sense that there were no fringe shifts. Thus, one is entitled to claim that the Kennedy–Thorndike experiment failed to produce the kind of positive effect whose occurrence would have served to *confirm* the Lorentz–Fitzgerald hypothesis. But it would be an *error* to suppose that the *non*-obtaining of this particular kind of confirmation suffices to prove that the Lorentz–Fitzgerald hypothesis was *falsified* by the null result of the Kennedy–Thorndike experiment! For we shall now see that the adjunction of the *further* auxiliary hypothesis of time dilation to the Lorentz–Fitzgerald hypothesis does enable the thus *doubly* amended aether theory to explain the null outcome of the Kennedy–Thorndike experiment while upholding the Lorentz–Fitzgerald hypothesis. And it will then become apparent that the justification for rejecting the Lorentz–Fitzgerald hypothesis along with the doubly amended aether theory depends on having philosophical reasons for accepting Einstein's rival theory of special relativity to the exclusion of the doubly amended aether theory ([1964], p. 1411; reprinted in ch. 21 of Grünbaum [1973], pp. 722–3).

Thus, in my 1964 article, I emphatically *denied* that the null result of the Kennedy–Thorndike experiment refutes the *LFC* by *explaining* why it does not. Yet eleven years later, Leplin ([1975], p. 314, n. 16, p. 315 and n. 20 on p. 315) cited that same article and asserted (p. 315) that in this very article, I had 'argued that the null result of the Kennedy–Thorndike experiment of 1931 constitutes a refutation of the contraction hypothesis'. And earlier, Leplin ([1972], pp. 486–7 and 490) had likewise cited my 1964 article when speaking (p. 486) of 'Grünbaum's argument to the effect that a negative result in the Kennedy–Thorndike experiment refutes the Lorentz contraction hypothesis'. The kernel of truth in these anachronistic misattributions is that in my ([1959]), in my ensuing exchange with Herbert Dingle in this *Journal* (Grünbaum [1960], [1961]), and elsewhere *before* 1964 I had indeed given such an argument, a fact noted by Leplin ([1975], p. 315, n. 20).

When pointing out why the *LFC* is not *ad hoc* (*c*) in my 1964 article, I had offered an explication of Einstein's different *ad hoc* charge against the *LFC* as follows:

> Since Einstein thus forsook the Lorentz–Fitzgerald version of the aether theory as *ad hoc* in his quest for a new rival theory, I deem it reasonable to take Einstein's rejection of the contraction hypothesis as '*ad hoc*' in this context to be tantamount to the following two-fold claim: (*i*) Prior to the *E*-result of the Michelson–Morley experiment, no other kind of experiment had an outcome

> providing support for the contraction hypothesis. (*ii*) It is to be expected that the Lorentz–Fitzgerald modification of the aether theory will *not* be confirmed by subsequent tests of a kind *different* from the Michelson–Morley experiment. And this conjecture serves as a reason for not accepting the Lorentz–Fitzgerald contraction hypothesis as an explanation of the null result of the Michelson–Morley experiment. Thus I take Einstein's rejection of the contraction hypothesis as *ad hoc* not only to allow but to assert that this hypothesis is indeed 'independently testable', that is, testable by experiments of a kind different from the Michelson–Morley experiment. For I construe Einstein's pejorative usage of the term '*ad hoc*' to refer to the posited fact that though independently testable, the contraction hypothesis would fail to secure subsequent independent experimental confirmation *as against the claims of a new rival theory* ([1964], p. 1409; reprinted in my [1973], pp. 716–7).

Leplin ([1975], p. 321 and n. 34) disregarded the last sentence of this statement when misinterpreting the earlier exegetical sentence in it which says that 'It is to be expected that the Lorentz–Fitzgerald modification of the aether theory will *not* be confirmed by subsequent tests of a kind *different* from the Michelson–Morley experiment'. The intended reading of the latter sentence is, of course, that *conjecturally* there are *some* independent tests which will fail to confirm the *LFC* while confirming the envisioned *STR*, *not* that *every* subsequent test will disconfirm the *LFC*. For my clarifying assertion that the *LFC* 'would fail to secure subsequent independent experimental confirmation *as against the claims of a new rival theory*' fully allows that the *LFC* be confirmed in *some* subsequent tests. But the point is that there would be *other* tests which are expected to disconfirm the *LFC* while confirming the new rival theory (*STR*).

Furthermore, it is elementary that every false theory has *some* true entailments. Hence, even if there already had been sufficient grounds at the time of Einstein's *ad hoc* charge for deeming the *LFC* to be false, I would hardly have considered *these* grounds sufficient for what Leplin ([1975], p. 321) calls 'the commitment to its [the *LFC*'s] disconfirmation by all possible experimental tests'. And yet Leplin (*ibid.*, n. 34) misattributes to me the very strong untutored claim 'that if an hypothesis *H* is introduced *ad hoc* [in the Einstein usage of which I had proposed an exegesis] as a response to an experiment *E*, then all experiments of a "different kind" from *E* which count as tests of *H* will disconfirm it' (*ibid.*, p. 321). Moreover, Leplin ([1975], p. 315, n. 21) says that in the 1964 paper 'Grünbaum limits the "*ad hocness*" of the [*LFC*] hypothesis to its allegedly illegitimate use in retaining the electron theory once the alternative of relativity was available'. If he is right about that, how can he consistently adduce just that earlier account of mine—as he does (*ibid.*, p. 315)—to illustrate the following complaint of his: 'This has lent impetus to the alternative view that no decision between the theories during their period of competition [n. omitted] could be rationally supported'?

Popper ([1959*b*] and [1973], p. 51, n. *1) accepted my demonstration that the *LFC* had never been *ad hoc* (*c*). But the details of his acceptance call for a critical appraisal which I shall give in section 3. For now I wish to illustrate the utility of the distinctions I have drawn between the 3 kinds of *ad hoc* auxiliaries by pointing out how preoccupation with one member of this hierarchical family to the neglect of the others prompted G. Holton ([1969], pp. 179–80) and me (Grünbaum [1973], Appendix §§34–5, pp. 834–5) to argue at cross purposes with one another in regard to the soundness of an *ad hoc* charge against the *LFC*.

Holton was rightly concerned to note the following point neglected by me: *When it was first propounded*, the *LFC* was *ad hoc* (*a*). But he goes on to say the following: 'Whether or not epistemological analysis can establish other, perhaps "largely unnoticed", senses for the phrase ["*ad hoc*"] remains to be seen. C. G. Hempel has shown clearly that "there is, in fact, no precise criterion for *ad hoc* hypotheses"' (*ibid.*, p. 180). Holton's cautionary and sceptical remark here is a response to my two-fold contention as follows: (i) Since the *LFC* has *never* been *ad hoc* (*c*), anyone who indicted it as *thus ad hoc* was mistaken, and (ii) the *LFC* had ceased to be even *ad hoc* (*b*) at least since the time when Kennedy and Thorndike designed the optical experiment which they performed in 1932. But Holton overlooks that Hempel's *caveat* cannot serve Holton's intended purpose of upholding his own *ad hoc* (*a*) charge while impugning my stated two-fold contention: As my account of Hempel's reasoning in 3*b* below will show, Hempel's doubts affect the precision of the (*a*) sense of '*ad hoc*', whose historical relevance Holton was concerned to stress in criticism of me, no less than the precision of the (*b*) and (*c*) senses of '*ad hoc*', which I had invoked. Hempel's doubts are a double-edged sword, because (1) they pertain to the feasibility of giving a characterisation of the concept of *independent* consequence in purely logical as distinct from pragmatic terms, and (2) this concept is ingredient in the (*a*) sense of '*ad hoc*' no less than in the (*b*) and (*c*) senses.

Finally a comment on the moral of a criticism of Zahar by Schaffner. Just as my definition of the *ad hoc* (*c*) attribute of an auxiliary involves a claim concerning the membership of the entire class $LC(T_2)$ rather than of only $KC(T_2,t)$, so also the '*ad hoc*$_1$' notion of Lakatos and Zahar ([1973], p. 101) involves such a strong claim in its *definiens*, where 'novel' consequences in the latter correspond to 'independent' consequences in the former. Thus, as also illustrated by Popper's cited *ad hoc* (*c*) charge against the *LFC*, claims pertaining to $LC(T_2)$ rather than only to $KC(T_2,t)$ have played a role in the history of the *philosophy* of science. But Schaffner ([1974], pp. 67–8) has contended in effect that $KC(T_2,t)$ rather than $LC(T_2)$ is the operative reference class for *ad hoc* charges levelled by working scientists. And hence Schaffner argued (*ibid.*, p. 68) that 'Zahar's first sense of *ad hoc*, in which a theory is *ad hoc* if "it has no novel consequences as compared with its predecessor", is a logical dream'. In the light of Schaffner's observation, I am inclined to think that the *ad hoc* (*b*) concept has greater relevance than the *ad hoc* (*c*) concept to episodes in the history of empirical science itself, as distinct from the history of the philosophy of science. This recognition and the hierarchical logical relations within my family of *ad hoc* concepts as well as the latter's temporal relativisation meet some objections raised by Leplin ([1975], pp. 313–16) against certain influential prior philosophical accounts of *ad hocness.*

3 IS THERE A PURELY LOGICAL, NON-PRAGMATIC CHARACTERISATION OF *ad hoc* AUXILIARY HYPOTHESES?

(*a*) *Critique of Popper's Falsificationist Account of Legitimating* ad hoc *Auxiliaries*

In response to my aforementioned demonstration that the *LFC* was never *ad hoc* (*c*), Popper wrote in this *Journal*:

> Professor Grünbaum's most lucid explanation of the Kennedy–Thorndike experiment (1932) is a very welcome contribution to the discussion of the problem whether or not the introduction of a new hypothesis is to be regarded as a potential advance. In the passage of my book referred to by Professor Grünbaum I said that the goodness of an auxiliary hypothesis would depend on whether it is independently testable, or perhaps merely rescuing a refuted theory. Professor Grünbaum points out, rightly, that I have been guilty of uncritically accepting (or possibly even spreading) the myth that the Lorentz–Fitzgerald contraction hypothesis was not testable. Professor Grünbaum's correction shows that this hypothesis was testable and thus not *ad hoc* to the degree I believed. Accordingly *it was an advance.* But it was, of course, more *ad hoc* than special relativity. In other words, we have here an excellent example of 'degrees of *ad-hocness*' and of one of the main theses of my book—that *degrees* of *ad-hocness* are related (inversely) to degrees of testability and significance ([1959*b*]).

In the fifth German edition of his *Logik der Forschung* ([1973], p. 51), Popper makes clear that the relative degrees of *testability* which are to specify inversely the comparative degrees of *ad hocness* are to be understood as comparative degrees of *falsifiability*, just as one would expect from him. For there he appended the following footnote to the reprinted original

German assertion of his *ad hoc* (*c*) charge against the *LFC* (translation is mine):

> *This is false:* the contraction hypothesis *does* have falsifiable consequences, as A. Grünbaum explains in the *British Journal for the Philosophy of Science*, **10** (1959), pp. 48–50. (But it is naturally testable to a lesser extent than the special theory of relativity and is therefore an example of the fact that there are *degrees of ad hocness*.) (Italics in original.)

We now need to appraise critically Popper's claim here that the *LFC* provides an excellent example of degrees of *ad hocness as construed* by his major *LSD* thesis that degrees of *ad hocness* are related inversely to degrees of falsifiability. In order to give this appraisal, we first need to recall two things as follows: (*i*) Popper's exact falsificationist criterion for the admissibility of theory-modification by an auxiliary hypothesis, and (*ii*) my proof elsewhere in this *Journal* (Grünbaum [1976*a*], pp. 6–9) that if comparative degree of falsifiability is understood in terms of the proper inclusion relation between specified kinds of content classes, then Popper's criterion of admissibility is not applicable even to one single case.

In his *LSD* ([1959*a*], section 20, pp 82–3) Popper writes:

> As regards *auxiliary hypotheses* [introduced to accommodate instances which refute the *original* form of the theory] we decide to lay down the rule that only those are acceptable whose introduction does not diminish the degree of falsifiability or testability of the system in question, but, on the contrary, increases it. (How degrees of falsifiability are to be estimated will be explained in sections 31 to 40.)

Upon turning to his sections 31–40 (pp. 112–35), we find him telling us (in sections 32, 33 and 38) that the degrees of falsifiability of two theories may be comparable not only *via* subset relations among their Tarskian content-classes but also by means of their '*dimensions*', *i.e.* by means of 'the degree of complexity of their simplest falsifiers' ([1974], Book II, p. 1047). And he then adds significantly ([1959*a*], p. 130) that there will be cases in which neither or only one of the two methods of comparison of degrees of falsifiability is applicable. But he does not tell us what to do when neither is applicable. In any case, it is clear that in his *LSD*, he intends the achievement of *increase of content* to qualify as *sufficient* to legitimate the postulation of a modifying auxiliary hypothesis without asserting that it is also *necessary*, although he does require an increase of the *degree of falsifiability* as certified by at least one of his two criteria. Yet in his 'Intellectual Autobiography', he seems to be saying that *increase of content* is both necessary and sufficient to legitimate a modifying auxiliary. For there he writes:

> in *Logik der Forschung* . . . I introduced *degrees of testability*, and these turned out to be closely related to (degrees of) *content*, and surprisingly fertile: increase of content became the criterion for whether we should, or should not, tentatively adopt an auxiliary hypothesis ([1974], p. 32).

In an earlier issue of this *Journal*, I have demonstrated the following result (Grünbaum [1976*a*], pp. 6–9): If the non-metrical content of a theory T is construed as either the Tarskian logical content $LC(T)$ or as Popper's informative content, or yet as Popper's empirical content (which is a species of his informative content), then no auxiliary H can *ever* generate *a rescuing* T_2 *of greater content*! And thus the desideratum that there be an increase of *any such* non-metrical content is a wholly *useless* and *vacuous* sufficient condition for compliance with Popper's rule of admissibility. Furthermore, if *such* content-increase were also held to be a necessary condition for the admissibility of H, then *no* modifying auxiliary could *ever* satisfy it. Moreover, there (*ibid.*, p. 6) I also inquired whether the construal of the non-metrical content of a theory T as its *question*-content $QC(T)$ can make Popper's criterion of content-increase non-vacuously applicable to the admissibility of rescuing auxiliaries after all. In particular, I considered whether the postulation of Neptune—which Popper explicitly deemed *legitimate* ([1974], Book I, p. 32)—can be held to comply with Popper's rule of admissibility in virtue of having generated a planetary theory of increased *question*-content. And I concluded (*ibid.*, p. 22) that it is at best doubtful whether Popper can rely on *question*-content to legitimate the introduction of the Neptune hypothesis and of other *prima facie* legitimate auxiliaries in accord with his general canon of content-increase. Furthermore, I noted that it is at best moot whether in any of the known cases of theory-rescue by a modifying auxiliary H, Popper's *dimension method* for comparing degrees of falsifiability is capable of assessing compliance with his stated LSD requirement for the admissibility of H. Combining these several considerations, I then reached the following overall conclusion (*ibid.*, p. 23): 'Thus, it is at best unclear whether Popper's criterion of admissibility for auxiliaries has any significant applications at all to concrete cases'.

Remaining mindful of this sober result, let us ask whether Popper can sustain his thesis that the LFC provides an excellent example of degrees of *ad hocness*, when these are construed as inversely related to degrees of falsifiability. When Popper tells us that degrees of *ad hocness* are related inversely to degrees of falsifiability, we are to understand that the former correspond to comparative degrees of *non*-acceptability while the latter furnish a comparative criterion as to whether one theory constitutes an *advance* over another. And for Popper ([1959*a*], pp. 82–3;

[1959*b*]) the proposed successor B of a given theory A is *maximally* and unacceptably *ad hoc* iff B's degree of falsifiability is not higher than A's. Thus, by demanding that T_2 have a *higher* degree of falsifiability than T_1, Popper's criterion for the admissibility of an auxiliary is not content with the lesser demand that T_2 merely *not* be less falsifiable than T_1: Mere avoidance of lesser falsifiability would have been a very weak requirement indeed, if only because it would always be satisfied when T_1 and T_2 just cannot be rank-ordered at all with respect to degree of falsifiability.

I now wish to exhibit the ill-foundedness of Popper's above claim that the *LFC* provides 'an excellent example of "degrees of *ad hocness*"'.

Let '$<_f$' denote the relation of actually being less falsifiable by at least one of Popper's avowed standards for degrees of falsifiability. And assume that '$>_f$' denotes the converse relation of greater Popperian falsifiability. In this context, T_1 is the original aether theory, while T_2 is the modified version thereof in which the *LFC* is ingredient. According to Popper, my demonstration that the *LFC* was *not ad hoc* (*c*) suffices to establish his claim that $T_1 <_f T_2$. And he can then infer as well that T_2 was an advance over T_1 by his methodological requirement. But his initial inference is unsound: His claim that $T_2 >_f T_1$ surely does not follow from the mere fact that T_2 does have an independent falsifiable consequence, *i.e.* that the *LFC* is not *ad hoc* (*c*)! For while the *LFC* has such a consequence, this fully *allows* the following: *Overall,* T_1 and T_2 are incomparable (not commensurable) with respect to *each* of Popper's avowed standards for rank-ordering theories according to $<_f$ or $>_f$. The latter incomparability is *illustrated* by the fact that T_2 *cannot* exceed T_1 in regard to Tarskian logical content *LC*, Popperian informative content *IC*, or Popperian empirical content *EC*. Such content incomparability obtains (*cf.* Grünbaum [1976*a*], pp. 6–9) because T_1 and T_2 are logically incompatible and yield contrary empirical predictions for the outcome of the Michelson–Morley experiment. Thus, there are no increases of *LC*, *IC* or *EC* which could possibly furnish a sufficient condition for T_2 to be more falsifiable than T_1. Nor has Popper given any grounds at all for concluding that the required sufficient condition is provided by either his question-content criterion or his dimension-of-a-theory criterion. Hence his assertion of $T_1 <_f T_2$ is ill-founded. The triplet of theories T_1, T_2 and *STR* are pairwise logically incompatible as shown by the pairwise incompatible predictions they make for the outcome of the Kennedy-Thorndike experiment. And for quite analogous reasons, Popper's further falsifiability rankings $T_2 <_f STR$, and $T_1 <_f STR$ are unfounded no less than his claim that $T_1 <_f T_2$. Incidentally, the transition from T_2 to the *STR* would seem to be more a matter of wholesale theory-replacement than of mere theory-*emendation* by an auxiliary.

Thus we see that Popper's *LSD* criterion for the admissibility of an auxiliary H had demanded that $T_1 <_f T_2$ but that the latter falsifiability-ranking is not assured on the strength of H's independent testability. Hence, in Popper's reply ([1959*b*]) to me, it was misleading if not outright fallacious on his part to have put the following gloss on the passage of his *LSD* to which I had referred when taking issue with his original *ad hoc* (*c*) charge against the *LFC*: 'I [Popper] said that the goodness of an auxiliary hypothesis would depend on whether it is independently testable, or perhaps merely rescuing a refuted theory'. This statement of Popper's ([1959*b*]) is at best misleading in its context, because he then trades on it as follows: He then fallaciously treats the fact that the *LFC* is *not ad hoc* (*c*) as tantamount to compliance with his *LSD* requirement that $T_2 >_f T_1$. Specifically, he commits this fallacy explicitly when using the term 'thus' in his ([1959*b*]) statement that 'Professor Grünbaum's correction shows that this hypothesis [*LFC*] was testable and thus not *ad hoc* to the degree I believed'.

I must conclude that, contrary to Popper's allegation, his version of a falsificationist criterion for the admissibility of auxiliaries and, more generally, of theory-replacements does not vindicate the transition from T_1 to T_2 and thence to the *STR*. For he has provided no grounds for his claim that the requisite conditions

$$T_1 <_f T_2 <_f STR$$

are actually fulfilled with respect to at least one of his avowed standards of greater falsifiability. Nor do I see that he can give the required falsificationist vindication of the transition from the planetary theory τ_1 to τ_2 or of other episodes of theory-modification in the history of science which he avowedly wishes to countenance as having constituted scientific advances. In this connection, I refer the reader to my detailed examination ([1976*a*], section 3) of Popper's attempt to depict the transition from Newton's to Einstein's theory of gravitation as an instance of increase in the degree of falsifiability (erotetic content). I also refer to my ([1976*b*] and [1976*c*]).

Using a radically different approach, which would not be countenanced by Popper because it is Bayesian, Schaffner ([1974], pp. 67–73) also favours a notion of *ad hocness* which allows for 'more or less seriously *ad hoc* hypotheses'. And Schaffner points out (private communication) that this kind of notion is being employed in current microbiological arguments concerning the merits of theories of the generation of antibody diversity. But Schaffner's ideas are still too programmatic to permit detailed assessment now. Like comments apply to some suggestions which Wesley Salmon has made along similarly Bayesian lines.

(*b*) *Hempel's Thesis That 'There is, in fact, no precise criterion for* ad hoc *hypotheses'*

Before turning to an account of Hempel's reasons for his stated negative thesis, I now wish to justify a claim I made, early in section 2, concerning the characteristics of an intuitively independent consequence. There I said that if there are to be any auxiliaries at all which qualify as *ad hoc* on the strength of lacking any ***independent*** testable consequences, then the following attributes do ***not*** jointly suffice as the defining characteristics of such consequences K: $K \notin LC(T_1)$, $K \notin LC(E)$, and $K \not\leftrightarrow H$. I can now show that *every* auxiliary H has consequences K—in the context of the corresponding T_2—which possess these three attributes. It will then follow that if their possession were taken as sufficient for the 'independence' of a given testable consequence, then *no* H could be *ad hoc* (*b*), let alone *ad hoc* (*c*). And even such intuitively *ad hoc* (*a*) auxiliaries as the Neptune hypothesis could then generally not qualify as such even initially.

Let 'A' denote a testable proposition such that $A \in LC(T_1) \cap LC(T_2)$ but $A \notin LC(E)$. There will always be such statements A which are *known* to have these properties. And often, at least one of them will already have been tested at the time t_0 when H is first propounded. Therefore, there are always testable conjunctions $A \cdot E$ such that $A \cdot E \in KC(T_2,t_0)$, where A does *not* belong to the consequence class of E but *is* a consequence of both T_1 and T_2. However, any such testable conjunction $A \cdot E$ is seen to have the three attributes which had been tentatively deemed sufficient for being an 'independent' consequence of T_2. And the possession of these three attributes by $A \cdot E$ would be known to the scientific community from t_0 onward. Hence if the three stated attributes were indeed deemed sufficient for 'independence', then no H could possibly qualify as either *ad hoc* (*b*) or *ad hoc* (*c*) at any time. Moreover, not only E but typically at least one statement A of the above kind will already have been tested successfully at the time t_0: For example, in the case of our two planetary theories τ_1 and τ_2, their common logical constituent 'there are at least six *intra*-Uranian planets (besides the Asteroids)' had this property when Neptune was first postulated. And thus if the three tentatively sufficient attributes of an 'independent consequence' were taken to be its defining characteristics, then the *already tested* conjunction $A \cdot E$ would qualify as 'independent', thereby disqualifying the Neptune hypothesis from being *ad hoc* (*a*) at the time t_0. And yet, intuitively this hypothesis did so qualify. *Mutatis mutandis*, the same can be said concerning Leverrier's postulation of Vulcan.

Laudan suggested that in order to exclude $A \cdot E$ from the set S of independent testable consequences, I should consider strengthening the three stated requirements for an independent consequence by demanding additionally that any such consequence K be atomic rather than molecular. But this proposed further necessary condition would be much too strong for at least the following reason:

Physical theories T_2 entail testable molecular statements of the conditional form $I \rightarrow F$, where '$\rightarrow$' denotes the entailment relation, 'I' denotes a certain set of initial conditions realised at the time t, and 'F' represents an outcome at a time $t+\Delta t$ which is predicted by T_2 under the conditions I. In short, since $(T_2 \cdot I) \rightarrow F$, it follows that $T_2 \rightarrow (I \rightarrow F)$. But given that T_2 thus entails testable non-atomic statements of the form $I \rightarrow F$, we would surely not wish to disqualify *all such* molecular statements from being 'independent' consequences of the auxiliary. For example, such an exclusion would disqualify the *optical* implications of the Neptune hypothesis for telescopic observations. Hence we cannot make it a necessary condition for $K \in S$ that K be atomic.

Suppose that we *try* to deal with the need to assure that $A \cdot E \notin S$ by imposing the following further exclusions from S along with a *caveat*: (*i*) If $K \in S$, then K may not be a conjunction of any of the statements already excluded from S by the prior necessary conditions for independence. (*ii*) If $K \in S$, then $K \notin LC(A \cdot E)$, where the *caveat* is that we do *not* countenance that *every* consequence of *every* non-independent conjunction is itself non-independent. The reason for thus *not* generalising the further exclusion (*ii*) is that countenancing the generalisation would have the worst possible absurd result as follows: T_2 could have no independent testable consequences whatever. This would be so, because assumedly T_2 can be expressed as a conjunction $T_2 = T_0 \cdot H$, where T_0 is also a consequence of T_1, while $(T_0 \cdot H) \notin S$ by our prior exclusions.

Let us suppose that the stated two-fold further requirement, as qualified by the above *caveat*, is not too strong. Even then, could we venture the claim that the conjunction of our explicit formal requirements for being an intuitively independent testable consequence of H (*i.e.* of T_2) captures the latter's defining characteristics? The answer to this question brings us to Hempel's thesis.

Hempel has contended that an analysis of the construal of the observational consequence E of T_2 poses formidable obstacles for the success of the attempted purely logical explication. Hempel had privately communicated his doubts to me for inclusion in my [1964], and my prior statement of them can be found there (pp. 1410–12) as well as in (Grünbaum [1973], pp. 719–20). Here I wish to give a less schematic, more illustrative statement of them by reference to the null result of the Michelson-Morley experiment, with the understanding that this statement is only my own gloss.

Let us be mindful of our earlier assertion that theories T_2 entail observational statements of the form $I \rightarrow F$. And let us suppose that in order to make E specific, we construe it to cover an infinite class of observation statements which differ only in the places and times to which they pertain. We know that Michelson and Morley did their experiment with their interferometer on at least one occasion by employing a light path of 11 meters, while D. C. Miller did such an interferometric experiment so that the light path had the much larger value of 64 meters (Jenkins and White [1950], p. 398). Now, in accord with our logical restriction on the meaning of 'E', let 'E' denote the consequence of the Lorentz–Fitzgerald theory T_2 that at any place or time, the 11 meter light path experiment yields a null result. Then the assertion F of a null result which T_2 makes for the outcome of the 64 meter experiment must count as an observational consequence *different* from E which satisfies all our tentative conditions for being 'independent'. But if F counts as an independent consequence of the *LFC* auxiliary, then we obtain the following counter-intuitive, unacceptable conclusions: (*i*) The LFC could *never* have been *ad hoc* (*b*), since it had, of course, been known from the outset that the modified, Lorentz–Fitzgerald aether theory T_2 made a prediction of the outcome of a 64 meter Michelson–Morley ('*M–M*') experiment no less than of an 11 meter one. (*ii*) After the length of the light path had been varied for the first time in a repetition of the *M–M* experiment, the *LFC* was not even any longer *ad hoc* (*a*).

Clearly, one can only scoff at so cheap and easy an exemption of the *LFC* from being *ad hoc* even in a descriptive, let alone a pejorative sense: Such a riposte would have done nothing at all to allay the misgivings against the *LFC* which were felt by Lorentz himself and by Poincaré, although they took the *LFC* seriously and were even willing to 'live with it'—at least for a while—because of its other presumed virtues. On the other hand, suppose that in order at least to *allow* the *LFC* to have been *ad hoc* (*b*) at some stage, one endeavours to specify *logical* criteria which would justify the following intuitive judgment: The 11 meter and 64 meter *M–M* experiments are (instances of) subvarieties of '*one single kind* of experiment' which tested a *single* kind of observational consequence *E*. How devise such purely logical—as distinct from pragmatic—criteria so that they would *also* underwrite the following *further* co-legitimate intuitive judgment: The Kennedy–Thorndike and *M–M* experiments *are* significantly different kinds of tests of the *LFC* in virtue of respectively testing *E* and an *independent* observational consequence *G*?

More generally, the moral of Hempel's doubts seems to be that given the initial tentative logical circumscription of *E*, there will always be *more* than one kind of observational consequence—and correlatively more than one kind of experiment—by which any given *H* is *known* to be testable at the outset. Hence no *H* could ever qualify as *ad hoc* (*b*), let alone as *ad hoc* (*c*). On the other hand, once the initial logical circumscription of *E* is relaxed, there seems to be no way to avoid the equally unwelcome result that a *prima facie* independent consequence *G* can then be accommodated within the membership of a suitably enlarged *LC*(*E*). And such accommodation would then undermine any argument which had relied on the independence of *G* to establish that *H* is *not ad hoc* (*b*). These considerations can serve as grounds for Hempel's published assertion ([1966], p. 30) that 'There is, in fact, no precise criterion for *ad hoc* hypotheses'. And hence he suggests that the notion of an independent consequence be explicated in terms of the pragmatic concept of an observational consequence which is *significantly* or *interestingly* different from the *E*-result.

I hope that my own prefatory indications of the difficulties encountered in giving a formal definition of 'independent consequence' serve to explain why I am inclined to take Hempel's doubts very seriously until and unless shown otherwise.

4 IS THERE A CLEAR NORMATIVE RATIONALE FOR DEROGATING *ad hoc* HYPOTHESES METHODOLOGICALLY AS A COGNITIVE LIABILITY?

If an auxiliary possesses a *prima facie* only *descriptive* attribute of being *ad hoc*, are there also *automatically* grounds for derogating such a hypothesis as a cognitive liability? If not, under what circumstances should an *H* which is descriptively *ad hoc also* be held to be *pejoratively* so?

Let us deal with this question by reference to several different accounts of *ad hocness*, including Popper's, Leplin's and Laudan's.

Popper writes:

> The question 'What kind of explanation may be satisfactory?' thus leads to the reply: an explanation in terms of testable and falsifiable universal laws and initial conditions. And an explanation of this kind will be the more satisfactory the more highly testable these laws are and the better they have been tested. (This applies also to the initial conditions.)
>
> In this way, the conjecture that it is the aim of science to find satisfactory

> explanations leads us further to the idea of improving the degree of satisfactoriness of the explanations by improving their degree of testability, that is to say, by proceeding to better testable theories; which means proceeding to theories of ever richer content, of a higher degree of universality, and of higher degree of precision (footnote omitted). This, no doubt, is fully in keeping with the actual practice of the theoretical sciences ([1972], p. 193).

Thus it is Popper's conjecture that science aims at ever more falsifiable theories. Hence it is easy for him to provide a normative rationale for cognitively derogating at least those auxiliaries which count as ***maximally ad hoc*** for him (*cf.* section 3 (*a*) above) on his proposed falsificationist scale of *degrees* of *ad hocness*. For, as we saw in section 3 (*a*), he conceives of these degrees as inversely related to degrees of falsifiability. But in a given context of inquiry, he can also indict auxiliaries which are *less* than maximally *ad hoc*: For example, though Popper concluded that the *LFC* was not maximally *ad hoc*, he regarded the *STR* to be still more falsifiable than the theory which incorporated the *LFC*. And in the face of the availability of the unrefuted *STR*, he therefore judged the *LFC* to be relatively *ad hoc* in both a descriptive *and* pejorative sense.

But, as I argued in 3 (*a*), it is at best unclear whether Popper's falsificationist criterion of admissibility for auxiliaries has any significant applications at all to concrete cases. Hence it is at best undemonstrated that the falsifiability rankings enjoined by his methodology can serve to vindicate those episodes of theory-modification by auxiliaries or of theory-replacement which he avowedly wishes to countenance as having been progressive (*e.g.*, the postulation of Neptune).

By the same token, the injunction to eschew auxiliaries deserving of Popper's *pejorative* '*ad hoc*' charge can provide little actual methodological guidance, if almost any auxiliary turns out to be ***maximally ad hoc***, simply because the theory incorporating it is not overall demonstrably more falsifiable than its predecessor. No comparable disability afflicts the putative exhortation to avoid postulating *ad hoc* (*b*) or *ad hoc* (*a*) auxiliaries. It would seem that Kant's dictum 'ought implies can' has relevance to methodological norms no less than to other kinds of injunctions.

Can one provide a rationale for cognitively derogating any auxiliary just because it has one or more of our *prima facie* merely descriptive *ad hoc* attributes (*a*), (*b*) and (*c*)? Or must the *core* theory T_0, which T_2 retained from T_1 by asserting the conjunction $T_0 \cdot H$, itself be held defective to *justify* the following derogation of H: H is cognitively unavailing and only cosmetic *qua being conjoined with* T_0?

One response to these questions is implied by Leplin's account ([1975], pp. 336–7) of those properties of an auxiliary H which he claims to be individually necessary and jointly sufficient for H to be *ad hoc*. He arrived at these properties by abstracting and generalising the *ad hoc* features of the *LFC*. He then tested his generalisation (*ibid.*, pp. 343–4) against the features of Pauli's 1930 neutrino hypothesis, which was assumedly *ad hoc*. Hence I shall now examine the pertinent part of Leplin's account.

Leplin (*ibid.*, p. 325) defines some categories of assessment for a theory T with respect to a problem P belonging to the 'domain' of T. One such category—the 'non-fundamentality' of a theory—is defined by him as follows (*ibid.*, p. 325): 'P shows that T is ***non-fundamental*** if no satisfactory solution of P can be achieved without the rejection of propositions in T and their replacement by propositions inconsistent with T'. And speaking of an auxiliary H which is introduced into a theory T in response to an experimental result E, Leplin states the last of his necessary conditions for the *ad hocness* of H as follows:

> There are problems other than E confronting T which there is good reason to hold are connected with E in the following respects:
>
> (*a*) these problems together with E indicate that T is non-fundamental,
>
> (*b*) none of these problems including E can be satisfactorily solved unless this non-fundamentality is removed,
>
> (*c*) a satisfactory solution to any of these problems including E must contribute to the solution of the others (*ibid.*, p. 337).

He calls this purportedly universal requirement for *ad hocness* the 'condition of non-fundamentality' (*ibid.*, p. 331).

The second of his five necessary conditions for *ad hocness* reads as follows:

> *Condition of justification*
>
> If an hypothesis H is introduced into a theory T in response to an experimental result E, then if H is *ad hoc*, E is evidence for H but:
>
> 1. No available experimental results other than E are evidence for H.
> 2. H has no application to the domain of T apart from E.
> 3. H has no independent theoretical support (*ibid.*, p. 320; also p. 337).

Bear in mind that Leplin refers to these necessary conditions respectively as the 'condition of non-fundamentality' and the 'condition of justification'. It is important to note that Leplin emphasises problems ***other than E*** in his condition of non-fundamentality. Hence when it is satisfied, along with his other necessary conditions for *ad hocness*, the following situation prevails: The replacement of a constituent M of the original theory T_1 by a ***contrary*** auxiliary H does ***not*** remove the non-fundamentality which the retained core theory T_0 has with respect to problems ***other than E***. In order to support his inclusion of the condition of non-fundamentality in his analysis of *ad hocness* for the *special* case of the *LFC*, Leplin argues in favour of the following averral:

> The most prominent critics of the contraction hypothesis, Poincaré, Einstein, and Lorentz himself, objected to Lorentz's conception of the ether as well. These objections have an important bearing on allegations of '*ad hocness*' against the contraction hypothesis (*ibid.*, p. 327).

Leplin then concludes his reconstruction of the pertinent history by writing:

> Two major reasons for judging an hypothesis to be *ad hoc* are apparent in this reconstruction. First, the hypothesis treats the experimental anomaly it is introduced to remove as a problem of completeness, whereas there is good reason to hold that the anomaly actually constitutes a problem of fundamentality. Second, the hypothesis contributes to the solution of only one problem although there is good reason to hold that any satisfactory solution to this problem must also contribute to the solution of other important problems. And the reason for so holding is that these problems alike arise from the same basic inadequacy of the theory—an inadequacy to be diagnosed on the basis of them as one of non-fundamentality (*ibid.*, p. 330).

But in order to justify the retention of the condition of non-fundamentality as a necessary condition in his *general* definition of *ad hocness*, Leplin simply asserts the following:

> All actual cases which satisfy the condition of justification and in which the concept of '*ad hocness*' is used also satisfy the condition of non-fundamentality. In cases satisfying the condition of justification but not non-fundamentality, no issue of '*ad hoc*ness' is raised (*ibid.*, p. 335).

Before I comment on the role of Leplin's non-fundamentality condition in regarding the *LFC* in particular to be *pejoratively ad hoc*, I should point out that Leplin's inclusion of this condition in his proposed *general* single definition of an '*ad hoc*' hypothesis draws no support from the general definition of '*ad hoc*' given in Webster's unabridged *New International Dictionary* (2nd edition, 1958, p. 32). Insofar as this dictionary definition can be presumed to reflect traditional *scientific* usage, the absence of any allusion to non-fundamentality from its *definiens* may well not be insignificant. Webster's definition reads as follows: 'pertaining to, or for the sake of, this case alone;—said of a hypothesis or reason adduced to explain a phenomenon or class of phenomena, and having no other application by which its validity and significance can be tested'. There is no hint here of any *requirement* of non-fundamentality, although the dictionary definition clearly *allows* that an auxiliary which is '*ad hoc*' as defined also satisfies the condition of non-fundamentality in the manner of the *LFC*. Nor is there any such general requirement in the cases from the scientific literature which have been canvassed by Holton ([1969], p. 178).

Leplin adduces as a kind of 'severe test' of his proposed definition of '*ad hoc*' that Pauli's initially *ad hoc* neutrino auxiliary instantiates all of the properties which Leplin had abstracted and then generalised from the *LFC* case. Important as the *LFC* and neutrino auxiliaries are qua *ad hoc* hypotheses, how do these *two* cases *alone* justify the inclusion of the non-fundamentality condition in the *definiens* of an *ad hoc* concept for which Leplin claims *unique* relevance to the history of physics (*ibid.*, p. 345)? And how is my hierarchically-structured family of three *ad hoc* concepts—any one of which *allows* that, for a given *H*, there is non-fundamentality *as well*!—vulnerable to Leplin's allegation to the following effect (*ibid.*, p. 345): Philosophers of science who maintain that working scientists have used the term '*ad hoc*' in more than one sense—say in both my (*a*) and (*b*) senses—unwarrantedly depict scientists as vascillating 'inexplicably' among a variety of senses of '*ad hoc*'?

More importantly, it does not really matter whether scientists actually used the *words* '*ad hoc*' to refer to one or more of the sets of properties of auxiliaries which are singled out by my three senses of '*ad hoc*'. For I contend that auxiliaries having one or more of these sets of properties *deserve* to be singled out because of their epistemic roles in historically interesting episodes. And Leplin grants the *latter* contention. But he objects that the scientific *ad hoc* concept is *not* the one which is used to call attention to the admittedly important epistemic features singled out by one or more of my three senses of '*ad hoc*'. In support of this objection, Leplin points to a presumed lack of evidence that Einstein's criticism of the *LFC* as '*ad hoc*' had anything to do with *LFC*'s *empirical testability.* Yet, in his aforecited 'condition of justification' for *ad hocness*, Leplin ([1975], p. 320, item 2) himself requires that '*H* has no application to the domain of [the theory] *T* apart from *E*'. How is such lack of application not essentially tantamount to the kind of empirical testability deficiency codified by the following *ad hoc* requirement of mine: *Either H* does not even have any independently testable consequences *or* none are known as such *or* at least such consequences may not already have been tested? Furthermore, why must Einstein's '*ad hoc*' concept be held to be *paradigmatic*? And why suppose that the use of metascientific terms like '*ad hoc*' is *univocal* among all (leading) scientists?

Two auxiliary hypotheses from planetary astronomy will now illustrate that the weakest of my three *ad hoc* concepts renders an important epistemic feature of auxiliaries whose postulation need not necessarily be liable to Leplin's *censorious* non-fundamentality but may call for only a non-pejorative *caveat*. The two cases in point are the respective assumptions that there is an extra-Uranian planet (Grosser [1962], Jones [1947], and references given in Berry [1961], p. 371, and in Hanson [1963], p. 413) and that there are one or more intra-Mercurial planets (Newcomb [1895], [1884], ch. III, §2, pp. 292–4, and [1911]) auxiliaries about which leading astronomers like G. B. Airy and Simon Newcomb respectively were rightly uneasy, as we shall now see. In the case of the intra-Mercurial planet(s), none other than Leverrier himself was thus epistemically uneasy (Hanson [1962], p. 369, n. 13).

In 1834, T. J. Hussey wrote to Airy, who became Astronomer Royal at Greenwich in 1835, and proposed that the discrepancies between the theoretically calculated and 'observed' orbits of Uranus might be explained by the existence of a perturbing extra-Uranian planet (Grosser [1962], pp. 49 and 93). Airy, who was no 'dust-bowl empiricist' (Hanson [1963], p. 413), replied in part as follows:

> I have often thought of the irregularity of *Uranus*, and since the receipt of your letter have looked more carefully to it. It is a puzzling subject, but I give it as my opinion, without hesitation, that it is not yet in such a state as to give the smallest hope of making out the nature of any external action on the planet (*ibid.*, p. 93).

And in 1837, Airy dissuaded Bouvard from analysing the Uranian discrepancies or 'errors' for the effects of an exterior planet by conjecturing that they are due to mistakes in the perturbations attributed to the *interior* planets! Thus, Airy wrote:

> I cannot conjecture what is the cause of these errors, but I am inclined, in the first instance, to ascribe them to some error in the perturbations. . . . If it be the effect of any unseen body, it will be nearly impossible to ever find out its place (*id.*).

In the same vein, not only Bessel (Newcomb [1911*a*], p. 227) but of course Leverrier were all too aware of the fallibility of the assumptions made—*e.g.* about the *masses* of perturbing *intra*-Uranian planets—in the calculation of the Uranian orbit which is theoretically expected in the *absence* of any exterior planet. Indeed, Grant ([1966], Appendix, p. 609) provides a table, constructed by Benjamin Peirce, which contains 'the results of Le Verrier's attempt to account for the irregularities of the planet [Uranus] without supposing it to be influenced by any foreign cause of disturbance'. The warrant for caution in regard to the *intra*-Uranian part of the accepted planetary theory was borne out for the nineteenth century astronomical community by the following fact (Pannekoek [1961], p. 363): Only after it was realised that the *terrestrial* mass, assumed after Encke's solar parallax, was too small by 1/10 did planetary theory cease to be beset by anomalies exhibited by the nodes of Venus's orbit and by the perihelion of Mars.

Moreover, even by the time Leverrier wrote to Galle (September 1846) that 'it is impossible to satisfy the observations of Uranus without introducing the action of a new Planet thus far unknown' (Grosser [1962], p. 116), he had made the following provision in his Academy Memoir of only a few months earlier: An alteration in Newton's law of gravitation was for him 'a last resort to which [he] would not turn until all other potential causes for the discrepancies had been investigated and rejected' (*ibid.*, p. 100). Thus, in principle, he *allowed* that his auxiliary hypothesis of an exterior planet might conceivably be falsified. No wonder that he was impatient when he urgently requested Galle to undertake an optical test of his auxiliary after neither the French observational astronomers nor the British Astronomer Royal gave heed to doing so in the manner desired by Leverrier (Grosser [1962], pp. 102, 105, 111–12, 114–16). Even if Leverrier did not feel that optical confirmation was needed to remove any residual doubts of his own, it was presumably necessary to convince the astronomical community.

In 1859, about thirteen years *after* Adams's and Leverrier's independent postulation of the specific elements of the hypothetical extra-Uranian planet, Leverrier thought that he could repeat his prior postulational feat to account for the unexplained residuum in the advance of the perihelion of Mercury which was known even before the discovery of Neptune (Hanson [1962], p. 367): He first postulated only a single intra-Mercurial planet 'Vulcan' (*ibid.*, pp. 372–4) concerning which the astronomer S. Newcomb (1835–1909) wrote the following ([1895], vol. 8, p. 569):

> The planet has been looked for on many occasions, especially during total eclipses of the sun, and some astronomers have believed that they saw it. But it is now fairly well settled that the supposed planet has no real existence, so well settled, in fact, that the question no longer appears in astronomical literature.

A like observational fiasco was in store for Leverrier's modified auxiliary postulation of a *group* of intra-Mecurial planets at least some of which are large enough to be individually visible (Newcomb [1911], p. 155 and Hanson [1962], p. 374, section IV). It was then concluded 'that if the group exists the members must be so small as to be [individually] entirely invisible. But in this case they must be so numerous that they should be visible [collectively] as a diffused illumination on the sky after sunset. Such an illumination is shown by the zodiacal light' (*id.*). Yet the fact that the zodiacal light appears to be near the ecliptic rather than steeply inclined to it contributed to Newcomb's abandonment of his own 1895 explanation of Mercury perihelion motion (Clemence [1943]) by the attraction of a ring of intra-Mercurial planetoids and prompted him 'to prefer provisionally the hypothesis that the sun's gravitation is not exactly as the inverse square' (*id.*). Such an alteration of Newton's law had already been envisioned by Airy in 1846 (Grosser [1962], p. 48) and by a number of others before him (Hanson [1962], p. 376).

But Newcomb then rejected any modification of Newton's law of gravitation after all on the basis of the following impressive explanatory success of Newtonian perturbation theory: 'The demonstration by E. W. Brown that the motions of the moon's perigee is exactly accordant with the Newtonian law of gravitation, seems to preclude the possibility of any deviation from that law' (Newcomb [1911], p. 155). And while thus re-embracing Newton's law, Newcomb only *tentatively* endorsed Seeliger's improvement upon his own hypothesis of a ring of intra-Mercurial planets by deeming Seeliger's auxiliary 'the most probable one in the present state of knowledge' (*id.*).

But the crucial point *vis-à-vis* Leplin's insistence on non-fundamentality as a necessary condition for *ad hoc*ness is the following: Astronomers who had epistemic misgivings about one or another version of the hypothesis of intra-Mercurial planets—*e.g.* at the stage of wondering whether Vulcan would be found—did *not* seem to have *predicated* their epistemic uneasiness on *otherwise* already entertained doubts of non-fundamentality concerning Newton's theory of gravitation. Nor did they even necessarily seem to have *coupled* the latter doubts with the former. Yet the auxiliary hypothesis of intra-Mercurial planetoids had enough logical features in common with Pauli's early neutrino hypothesis—the latter of which Leplin *does* countenance as *bona fide ad hoc*—to have prompted Hanson ([1962], p. 370, n. 16) to characterise their 'comparison' as 'irresistible' (*cf.* also Hanson [1963], p. 414). Furthermore, it would seem that the mere widespread high confidence in Newton's theory of gravitation in the nineteenth century astronomical community did not assure a corresponding measure of confidence that the theoretical postulation and subsequent observational discovery of additional planets—rather than only *alternative* auxiliary hypotheses—would successfully resolve puzzles about the presumably known planetary orbits.

Thus we saw that at *some* stage, the *ad hoc* (*a*) postulation of an extra-Uranian planet had also twice evoked an attitude of epistemic *caveat* from no less a nineteenth-century astronomer than the Astronomer Royal at Greenwich.

So much for these particular historical episodes *within* physical science. Quite apart from these, my account accommodates or is at least germane to influential methodological debates in the history of the philosophy of science partly *because* it excludes pejorativeness from the *definiens* of *ad hocness*: As Larry Laudan has pointed out to me (privately), thinkers like J. S. Mill ([1887], Book III, ch. 14, §6, pp. 356–9) and Whewell debated whether hypotheses of which my *ad hoc* auxiliaries are a species should *or* should *not* be deemed methodologically objectionable.

These considerations contain only *some* of my reservations against Leplin's insistence on making non-fundamentality *quite generally* a necessary condition for *ad hocness*. But having so far at least indicated why I am reluctant to follow suit, I hasten to explain why I believe nonetheless that Leplin has rendered a valuable service by the articulation of his condition of non-fundamentality. I can do so by means of the following two assertions:

(*i*) There seems to be no good rationale for cognitively *derogating* an auxiliary *solely because* it has one *or* more of my—*prima facie* merely descriptive—*ad hoc* attributes (*a*), (*b*) and (*c*); indeed there is good reason for thinking that the *mere* possession of one or more of these *ad hoc* characteristics is insufficient to warrant derogation as artificial, *contrived* or arbitrary.

(*ii*) As in the *LFC* case, Leplin's condition of non-fundamentality does supply a rationale for cognitively *indicting* an auxiliary H as *merely cosmetic* when H is at least *ad hoc* (*a*) *and* is incorporated in the following kind of theory T_2: The *core* theory T_0, which T_2 rescued from the demise of T_1 by asserting the conjunction $T_0 \cdot H$, is *itself* separately suspected of *non*-viability. In *such* a case, the rationale for deeming H to be even pejoratively *ad hoc* is that H is cognitively *unavailing* qua being conjoined in T_2 with an avowedly non-viable core theory T_0!

I need to justify the first of these assertions. Within the framework of his own innovative problem-solving model of scientific growth, Laudan has well articulated the grounds which I myself adduce in support of this thesis (*i*). These grounds are *largely* applicable to my definitions of *ad hoc* (*a*), (*b*) and (*c*) even though Laudan gives a different definition of *ad hocness* which seeks chiefly to codify falsificationist accounts. He writes:

> *a theory is ad hoc if it is believed to figure essentially in the solution of all and only those empirical problems which were solved by, or refuting instances for an earlier theory.*
>
> Clumsy as it is, this characterisation of adhocness seems to do justice to some of the most sophisticated accounts of adhocness developed in the last decade. Assuming that adhocness is understood in this way, we are entitled to ask: *what is objectionable about it?* If some theory T_2 has solved more empirical problems than its predecessor—*even just one more*—then T_2 is clearly preferable

> to T_1, and, *ceteris paribus*, represents cognitive progress with respect to T_1 [footnote omitted]. However, we can go further than this to claim that the resort to *ad hoc* stratagems, as defined immediately above, is perfectly consistent with the general aim of increasing our problem solving capacities. Ad hoc modifications, *by their very definition*, are empirically progressive. . . .
>
> In urging that adhocness (so defined) is a cognitive virtue rather than a vice, I am clearly not implying that ad hoc theories are invariably better than non-ad hoc ones. My claim, rather, is that an ad hoc theory is preferable to its non-ad hoc predecessor (which was confronted with known anomalies). . . .
>
> But it might be argued that I have missed the point of the critics of *adhoc*ness. They might say, 'Yes, of course, T_2 is better than its *refuted* predecessor T_1; but the relevant comparison is between the ad hoc T_2 and some other theory T_n which is not ad hoc but still solves as many problems as T_2'. Einstein's special theory of relativity might exemplify T_n while the Lorentz-modified aether theory was T_2 [footnote omitted]. The obvious reply to such criticism is to ask why the admittedly *ad hoc* character of the Lorentz contraction constitutes a decisive handicap against it in comparing it with special relativity. If the empirical problem-solving capacities of the two theories are, so far as we can tell, equivalent, then they are (empirically) on a par; defenders of the view that the adhocness of T_2 makes it distinctly inferior to T_n must spell out why, in such cases, the comparable problem-solving abilities and equivalent degrees of empirical support can be thrown to the winds simply by stipulating that ad hoc theories are intrinsically otiose.
>
> What seems to lie behind many discussions of adhocness is a conviction—often present but rarely defended—that there is something suspicious about any change in a theory which is motivated by the desire to remove some anomaly. . . . We must remember that, as adhocness has been defined, any ad hoc change must *increase* rather than decrease the problem-solving capacity of the theory in question. Most of the obvious and trivial ways of eliminating anomalies—*e.g.*, arbitrarily restricting the boundary conditions, eliminating those postulates of the theory which entailed the anomaly (assuming they could be localised!), re-defining terms or correspondence rules—would generally result in *decreasing* the problem-solving effectiveness of a theory. Hence, such manoeuvres—which we might well wish to criticise [footnote omitted]—do not qualify as ad hoc. The detractors of adhocness have yet to show that the emendation of a theory to preserve its problem-solving capacity and to save it from an anomaly requires any less theoretical imagination or serendipity than the construction of a new theory from scratch. To the extent that these same detractors set an epistemic premium on theories which work the first time around, without any juggling or ad hoc adjustments, we are entitled to ask for the rationale for such a preference (Laudan [1977], ch. III, pp. 115–17).

Turning to my assertion (*ii*), note that there I was careful to say that H is 'at least *ad hoc* (*a*)'. Thus when non-fundamentality is likewise present, I clearly provide scope for the cognitive derogatability of an H which is *only ad hoc* (*a*) rather than *ad hoc* (*b*) or (*c*). Hence my present account enables me to handle the following kind of case which Leplin (*ibid.*, p. 316 and n. 24) posed as a challenge to prior characterisations of *ad hocness* in terms of testability: At a time when Lorentz knew his *LFC* to be independently testable via the Rayleigh–Brace experiments—so that then the *LFC* was no longer *ad hoc* (*b*) but only *ad hoc* (*a*)—his methodological misgivings concerning the hypothesis were not mitigated by this independent testability. Incidentally, the fact that the *LFC* was never *ad hoc* (*c*) and that there was a time after which it was only *ad hoc* (*a*) prompted my *not* defining a fourth sense of '*ad hoc*' which would be even stronger than *ad hoc* (*c*) by likewise incorporating Leplin's condition of non-fundamentality as a further necessary condition. Clearly, there is *no* time at which the *LFC* would have qualified as *ad hoc* in this putative still stronger sense, and I know of no other auxiliaries which would warrant its introduction.

It would seem, therefore, that my account is more flexible than Leplin's amid avoiding the following complaint of his against the 'elaborate evolution of qualifications and distinctions' which the 'connection between *ad hocness* and the falsifiability requirement has undergone':

> This vast proliferation is largely a response, as we shall see, to the discovery that, although clearly *ad hoc* in the judgment of the scientific community, the contraction hypothesis does not so qualify on falsificationist grounds. As such, these distinctions and refinements constitute something like a degenerating research programme, many of whose entries are patently *ad hoc* (*ibid.*, p. 314, n. 16).

I submit that the family of concepts which I defined by means of the (*a*), (*b*) and (*c*) senses of '*ad hoc*' can usefully render important epistemic and quasi-logical features of an historically relevant class of auxiliaries. And I use Leplin's condition of non-fundamentality to articulate a special condition under which the introduction of such *ad hoc* auxiliaries may warrantedly be regarded as *unavailing*. Their introduction may be so regarded under that condition NOT just because they are *ad hoc* in one or more of the senses (*a*), (*b*) and (*c*) but rather because *additionally* there are good reasons for suspecting T_0 of cognitive non-viability. Hence in that event T_2 *is derogated not because it introduced* H *but because it sought to rescue* T_0: The arrow of the *modus tollens* which the dramatic finding E directed at T_1 was already aiming at the heart T_0 of T_1 when H was first introduced. This state of affairs was presumably implicit in Einstein's derogation of the *LFC* as *ad hoc*. And my account thus accommodates Leplin's non-fundamentality *without* including it, however, among the necessary conditions in the *definiens* of *ad hocness*, as he does. Unlike Leplin, I do not thus include it, because (*i*) I deem it useful to single out a concept of *ad hocness* which is applicable to a wider class of epistemic cases, and (*ii*) I do not see why *all* of the logical features of Einstein's *ad hoc* charge against the *LFC* have to be paradigmatic for that concept. But I evidently do not deny that certain uses of the term '*ad hoc*' are intended to be derogatory: While granting these cognitive uses, I deny only that when *ad hocness* is indeed viewed as an unacceptable methodological liability, this pejorative assessment is necessarily always predicated on Leplin's non-fundamentality condition. Yet I emphasise that the fulfilment of the latter condition can supply an otherwise missing *justification* when an *ad hoc* auxiliary is rejected as illegitimate.

That the use of the '*ad hoc*' label by scientists does not possess the univocity claimed by Leplin would seem to be borne out by the following summary by Holton ([1969], p. 178) of usages which he has run across in the literature:

> The scientist who adopts somebody's hypothesis or creates his own for a specific purpose, 'in order to account' for a bothersome result or feature of the theory, regards it as ad hoc—not necessarily in a derogatory sense—*regardless* of its 'logical' status.... Thus we have found in the scientific literature characterisations of the following kinds for acceptable ad hoc hypotheses: 'not inconceivable', 'reasonable', 'plausible', 'fundamental', 'natural', 'appealing', 'elegant', 'likely', 'assumed a priori to get the desired results', 'auxiliary' or 'working hypothesis'. On the other hand, when an ad hoc hypothesis is rejected, we see it described in the following way: 'artificial', 'complex', 'contrived', 'implausible', 'bothersome', 'unreasonable', 'improbable', 'unlikely', 'unnecessary', 'ugly'. Sometimes 'ad hoc' itself is applied in the pejorative sense, and then it has such meanings.

Laudan ([1977], ch. III) discusses another interesting possibility for the warranted cognitive derogation of an auxiliary H as follows: There are *no* grounds for suspecting T_0 of non-fundamentality, and the suspicion of cognitive non-viability—which Laudan conceptualises as reduced problem-solving effectiveness—pertains only to the *conjunction* $T_0 \cdot H$.

University of Pittsburgh

REFERENCES

Agassi, J. [1975]: *Science in Flux.*

Berry, A. [1961]: *A Short History of Astronomy*, Dover Reprint.

Cantor, G. [1975]: 'The Edinburgh Phrenology Debate 1803–1828', *Annals of Science*, **32**, pp. 195 ff.

Clemence, G. M. [1943]: 'The motion of Mercury 1765–1937', *Astronomical Papers American Ephemeris*, **11**, pp. 1 ff.

Grant, T. [1966]: *History of Physical Astronomy*, Johnson Reprint.

Grosser, M. [1962]: *The Discovery of Neptune.*

Grünbaum, A. [1959]: 'The falsifiability of the Lorentz–Fitzgerald contraction hypothesis', *The British Journal for the Philosophy of Science*, **37**, pp. 48–50.

Grünbaum, A. [1960]: 'The falsifiability of the Lorentz–Fitzgerald contraction hypothesis: A rejoinder to Professor Dingle', *The British Journal for the Philosophy of Science*, **42**, pp. 143–5.

Grünbaum, A. [1961]: 'Professor Dingle on falsifiability: A second rejoinder', *The British Journal for the Philosophy of Science*, **46**, pp. 153–6.

Grünbaum, A. [1964]: 'The Bearing of Philosophy on the History of Science', *Science*, **143**, pp. 1406–12.

Grünbaum, A. [1973]: *Philosophical Problems of Space and Time*, 2nd edition.

Grünbaum, A. [1976*a*]: 'Can a theory answer more questions than one of its rivals?', *The British Journal for the Philosophy of Science*, **27**, pp. 1–23.

Grünbaum, A. [1976*b*]: 'Is the method of bold conjectures and attempted refutations *justifiably* the method of science?', *The British Journal for the Philosophy of Science*, **27**, pp. 105–36.

Grünbaum, A. [1976*c*]: 'Is falsifiability the touchstone of scientific rationality? Karl Popper versus inductivism', in *Essays in Memory of Imre Lakatos*, pp. 213–52.

Hanson, N. R. [1962]: 'Leverrier: The zenith and nadir of Newtonian mechanics', *Isis*, **53**, pp. 359–78.

Hanson, N. R. [1963]: Review of Grosser's [1962], *Isis*, **54**, pp. 413–14.

Hempel, C. G. [1966]: *Philosophy of Natural Science.*

Holton, G. [1969]: 'Einstein, Michelson, and the crucial experiment', *Isis*, **60**, pp. 133–97.

Jenkins, F. A. and White, H. E. [1950]: *Fundamentals of Optics*, 2nd edition.

Jones, H. S. [1947]: 'G. B. Airy and the discovery of Neptune', *Popular Astronomy*, **55**, pp. 312–16.

Laudan, L. [1977]: *Progress and Its Problems*, ch. III, 'Adhocness and the Evolution of Research Traditions'.

Leplin, J. [1972]: 'Contextual falsification and scientific methodology', *Philosophy of Science*, **39**, pp. 476–90.

Leplin, J. [1975]: 'The concept of an ad hoc hypothesis', *Studies in History and Philosophy of Science*, **5**, pp. 309–45.

Mill, J. S. [1887]: *A System of Logic*, 8th ed.

Newcomb, S. [1884]: *Popular Astronomy*, 5th ed., revised, ch. III, §2, pp. 292–4.

Newcomb, S. [1895]: 'Vulcan', *Johnson's Universal Cyclopedia*, **8**, p. 569.

Newcomb, S. [1911]: 'Mercury', *The Encyclopedia Britannica*, 11th ed., **18**, pp. 154–5.

Newcomb, S. [1911*a*]: 'Neptune', *The Encyclopedia Britannica*, 11th ed., **19**, p. 386.

Pannekoek, A. [1961]: *A History of Astronomy.*

Popper, K. R. [1959*a*]: *The Logic of Scientific Discovery.*

Popper, K. R. [1959*b*]: 'Testability and "ad-hocness" of the contraction hypothesis', *The British Journal for the Philosophy of Science*, **10**, p. 50.

Popper, K. R. [1972]: *Objective Knowledge.*

Popper, K. R. [1973]: *Logik der Forschung*, 5th German edition.

Popper, K. R. [1974]: 'Intellectual Autobiography' and 'Replies to my critics' in P. A. Schilpp (ed.): *The Philosophy of Karl Popper*, Library of Living Philosophers.

Schaffner, K. [1974]: 'Einstein versus Lorentz: Research programmes and the logic of theory evaluation', *The British Journal for the Philosophy of Science*, **25**, pp. 45–78.

Zahar, E. [1973]: 'Why did Einstein's programme supersede Lorentz's? (1)', *The British Journal for the Philosophy of Science*, **24**, pp. 95–123.

Brit. J. Phil. Sci. 27 (1976), 363–402 *Printed in Great Britain*

Discussions

VERISIMILITUDE REDEFLATED

In the same paper ([1974]) in which he recounted his dissatisfaction with Popper's theories of verisimilitude Professor Pavel Tichý gave the reader a glimpse of a new, to his mind logically acceptable, theory of distance from the truth. To me, however, the new theory seemed to have a grievous fault, and I accordingly appended to my own negative strictures [1974*a*] on Popper's theory what I thought was a clear refutation of the outlined proposal.

Tichý has now [1976] developed his theory in considerably greater detail, and in the midst of his exposition he offers a direct rebuttal of the criticisms I had made. This rebuttal I find singularly unconvincing, and even misleading. The developed theory too seems to have a lot more wrong with it than could have been anticipated from the earlier sketch. It is the purpose of this note to articulate these disagreements.[1]

The first section summarises, though only briefly (since there is a limit to how much repetition the *Journal* can bear), Tichý's earlier stated view and my criticism of it. I shall also derive an absurd consequence from Tichý's rejoinder. In the second section I shall discuss a recent criticism of Sir Karl Popper's. In the next two sections I turn to Tichý's fully developed theory, explain it as well as I can in brief compass, and show that it fails to satisfy Tichý's own intuitive requirements. To conclude I shall try to pinpoint what it is that separates my approach to the problem of verisimilitude from Tichý's.

1 We start with a language whose atomic sentences $p_0, p_1, \ldots$ are not any further analysed. Provided that these sentences are completely independent (that is, any selection of them can be true whilst the others are false), as would standardly be assumed, every complete consistent theory (constituent) of the language can be axiomatised by choosing for each i either p_i or $\neg p_i$. Of these constituents one will be *true*, the remainder *false*. Tichý argued ([1974], p. 159) that of two constituents one that contains all the true atomic sentences that the other does, and at least one more besides, is to be regarded as closer to the truth. This is a kind of dominance principle and, not surprisingly, it does not allow all constituents to be compared directly (one may contain p_0 and $\neg p_1$ while another contains $\neg p_0$ and p_1). Tichý, however, immediately proceeded to a universal method of comparison; indeed, it would be fair to say that he did not really isolate the dominance principle. The more general method is best restricted to languages where the number of atomic sentences is finite (though it does not have to be). It amounts simply to counting the number of false atomic sentences in each constituent. This method is clearly an extension of the dominance principle.

Tichý also gave (*loc. cit.*) an extension of this measure to all sentences of the language. I did not criticise this extension in my [1974*a*]. I shall turn to it in section 2.

---

[1] For helpful comments on an earlier draft I am indebted to Peter Mott, Professor I. Niiniluoto, Sir Karl Popper, Professor J. W. N. Watkins and John Worrall.

My criticism of Tichý's theory was based on the observation that his comparisons of verisimilitude were not invariant under translation into other (essentially equivalent) languages. In [1974*a*], pp. 175 f., I illustrated this with a concrete finite example based on Tichý's own discussion. (The condition of finiteness is redundant, as was in effect demonstrated in THEOREM 5 of my [1974*b*] (pp. 186 f.).) In his present paper Tichý calls this EXAMPLE A. Suppose that our language is limited to three atomic sentences, namely h, r, w. In terms of these we define the sentences m ($= h \leftrightarrow r$) and a ($= h \leftrightarrow w$); it is easily proved that h, m, a compose a completely independent triad too.[1] In [1974*a*] I asserted, originally without thinking the point to be controversial, that 'the theories h & m & a and h & r & w are equivalent theories'. Tichý unexpectedly denied this. He writes ([1976], p. 35):

> I, for one, find this last contention quite easy to deny. Two statements are equivalent if they say the same thing, *i.e.* if they have the same affirmative force. Now what is the affirmative force of a statement? It has been a commonplace at least from Carnap's *Meaning and Necessity* to identify this force with the *range* of that statement, *i.e.* with the class of possible states of affairs in which the statement holds true. ... What is a possible state of affairs? It is a distribution of truth-values through the class of atomic propositions of the language in hand, a function, that is, which takes the atomic propositions to the truth-values. ... Since $\{h, r, w\}$ and $\{h, m, a\}$ are clearly two distinct sets of propositions, no function defined on the former set can be identical with one defined on the latter.

It presumably follows from these remarks that if two languages do not have all their atomic sentences in common then no sentence of one will be translatable into any sentence of the other. Certainly this will be the case if they have no atomic sentences in common. Even Quine might question this. After all, it is well known that the theory of linear ordering can be formulated in two distinct ways (at least): in a language whose only primitive predicate is $\leqslant$, and in one with both $<$ and $=$ as primitives. Likewise, the theory of Boolean algebras can be written entirely in terms of an ordering relation $\leqslant$, or alternatively in terms of $+$, $\cdot$, $-$, and $=$. There are countless similar examples in abstract algebra. (One worth mentioning here is the equivalence, noted in my [1977*b*], between the theory of *distance* in Boolean algebras and the theory of *probability*.) Tichý's position, if valid, would compel us to look on these as so many different group theories, so many different lattice theories, and so forth. Even Feyerabend might frown at this.

I am quite happy to agree with Tichý that *two sentences in the same language have the same assertive power if and only if they hold in the same 'possible states of affairs'*. But it hardly follows from this that sentences *in different languages* may not have the same assertive power. And assuredly there is a problem of explaining under what conditions this obtains; it is undoubtedly not enough, as Tichý notes, to repeat the conditions that suffice in the case of a single language. In my original comments I acknowledged the existence of a problem by supplying references to two works where the problem had been discussed (Kanger [1968], Hanson [1961]; in my [1975*a*] I also mentioned Popper [1967]). Tichý appears not to have taken up these references. One would have thought that he would have welcomed an extension of the rather limited criterion that he credits to Carnap. But if he finds the suggestions of Kanger, say, unsatisfactory, there is no clue in his present paper why he does so. Instead he attaches himself firmly to a view that could only be at all plausible if there were an *absolute* (non-linguistic) distinction between atomic and molecular sentences; that is if, in the manner of Wittgenstein's *Tractatus* [1921], the question of which sentences are atomic is settled in some way by the structure of the world. Wittgenstein never argues for this view, and Black's attempt ([1964], p. 47) to provide an argument is far from satisfactory (see my [1977*c*]). On the other hand, arguments for treating the distinction as no more than relative were advanced by Popper in [1934] (pp. 127 f.), and are if anything strengthened by the type of example I have used against Tichý. What is more, Tichý himself acknowledges that 'Miller may well be right in maintaining that it would be sheer prejudice to treat *h*–*r*–*w*–ese as more fundamental than *h*–*m*–*a*–ese' (*loc. cit.*). Thus there seems to me to be little reason for taking Tichý's narrow theory of translatability seriously, and every reason for embracing a more comprehensive—and, I may say, more realistic—approach.

---

[1] The credit for noting that there is no unique completely independent set of sentences describing any state of affairs appears to belong to Black [1964], pp. 46 f. I was not aware of this when I wrote my [1974*a*] and [1974*b*]; otherwise I would have acknowledged Black's undoubted priority. Black's example, on the other hand, is not too elegant. For further comments see my [1977*c*].

2 All my criticisms of Tichý to date have been directed against the *lack of generality* of his definition of verisimilitude. Nothing in what I have said would prevent sentences in *some* languages from having degrees of verisimilitude as suggested by Tichý. Professor Popper has recently argued ([1976]), however, that Tichý's measures, extended to take in all sentences of the language and not solely the constituents, could *never* be appropriate. (A similar point has been made by Peter Mott in a private communication. It was Mott's letter that made me think more about the problems of this section.) I wish to make a few comments about this issue here.

Every consistent sentence of a finite sentential language can be expressed as a disjunction (perhaps with a single disjunct) of constituents—its disjunctive normal form. Tichý proposed ([1974], p. 159; see also [1976], pp. 26, 33) that the distance from the truth of a consistent sentence be identified with the average number of errors made per constituent of the sentence's disjunctive normal form. Thus to return to the earlier example, where the true constituent t was $h \,\&\, r \,\&\, w$, the sentence $h \,\&\, \neg r$, that is $(h \,\&\, \neg r \,\&\, w) \vee (h \,\&\, \neg r \,\&\, \neg w)$, makes one error in its first constituent and two in the second, so that its distance from the truth is 3/2. The tautology, on the other hand, makes twelve atomic errors in its eight constituents, so also has distance 3/2 from the truth.

Popper has correctly pointed out that this definition violates the following intuitively obvious condition: if a and b are *true* and $b \Vdash a$ then b is at least as close to the truth as a is.[1] The stronger (if there is one) of two true theories is not further from encompassing the whole truth. (In my [1977*a*] I call the condition that if $c \Vdash b \Vdash a$ then b is at least as close to c as a is the (weak) ***monotony condition*** for distance in a Boolean algebra.) For $(h \,\&\, r \,\&\, w) \vee (\neg h \,\&\, \neg r \,\&\, \neg w)$ has distance 3/2 from the truth, whilst many of its consequences are actually rather closer: for example, $(h \,\&\, r \,\&\, w) \vee (\neg h \,\&\, \neg r \,\&\, \neg w) \vee (h \,\&\, r \,\&\, \neg w)$ has distance 4/3 from the truth.

---

[1] I continue to use $b \Vdash a$ to mean that a is derivable from b. (See my [1974*a*], p. 170.) I use $b \vdash a$ to mean that a is derivable from b, but not conversely.

For convenience I shall in this section adapt some of Tichý's notation to my own use. I shall use $\triangle$, as he does, to represent distance, but I shall mean by $\triangle_n(b)$ ***the distance of b from the truth, computed for a language with n atomic sentences.*** (When n is irrelevant I shall omit it. The reason for not always doing so will emerge.)

It is not hard to see why Tichý's measure suffers from the defect noted by Popper and Mott. Suppose a true sentence a has m disjuncts in its disjunctive normal form, and these m disjuncts between them commit k atomic errors. Then $\triangle(a) = k/m$. A true sentence b obtained from a by subtracting ***one false disjunct*** will contain ***at least one error fewer*** (since the true disjunct must not be removed). In consequence, $\triangle(b) \leqslant (k-1)/(m-1)$. Under what conditions is $\triangle(b)$ guaranteed to be no greater than $\triangle(a)$? Simply: when $k \leqslant m$.

Thus the irregularity cannot arise if $k \leqslant m$; that is, if $\triangle(b) \leqslant 1$ for every true sentence b. (This is in fact the case for the measure $\triangle_2$.) As our calculations above are based on the assumption that ***each extra constituent*** and ***each extra error*** count ***one*** in the denominator and numerator respectively, our simplest move is to add a constant to the denominator to ensure that for true sentences it is never exceeded by the numerator. (It will clearly not do just to rescale $\triangle$; no ordering relations will be rectified by such a procedure.)

The weakest true sentence is the tautology, which I shall write ***taut***. We obviously need this sentence to be at least as far from the truth as any true sentence is. Thus it is sufficient to arrange that $\triangle(\textit{taut})$ is at most 1. In a language with n atomic sentences there are 2^n constituents. What we now need to compute is the total number of atomic errors these 2^n constituents commit. The number of constituents making r errors is $\binom{n}{r}$, so the total number of errors in all constituents is $\sum_{r=0}^{n} r.\binom{n}{r}$, which is equal to $n.2^{n-1}$. Thus the distance $\triangle_n(\textit{taut})$ of a tautology from the truth is, according to Tichý's measure, $n/2$. (Check: if $n = 3$ the distance is $12/8 = 3/2$.) Should we want the tautology to have distance exactly 1 from the truth we could define generally

$$\triangle'_n(a) = k/(m+2^{n-1}(n-2)),$$

where k and m are defined for a as before. We would have, then, $\triangle'_3(a) = k/(m+4)$. In the example worked above, $(h \,\&\, r \,\&\, w) \vee (\neg h \,\&\, \neg r \,\&\, \neg w)$ would have distance $3/6 = 1/2$ from the truth; its consequence $(h \,\&\, r \,\&\, w) \vee (\neg h \,\&\, \neg r \,\&\, \neg w) \vee (h \,\&\, r \,\&\, \neg w)$ would have distance 4/7 from the truth. The monotony condition is accordingly not breached.

Alternatively we could require that it is not the tautology that is constantly a distance 1 from the truth t, but the weakest false sentence $\neg t$. In a language with n generators the sentence $\neg t$ contains $2^n - 1$ constituents, but makes all the errors. Thus we would need to define

$$\triangle''_n(a) = k/(m+1+2^{n-1}(n-2)).$$

On this definition the tautology will always have a distance from the truth of less than unity.[1]

Each of these new definitions seems to me to be well within the spirit of Tichý's proposal and to be clearly better than Tichý's own definition. Nevertheless, neither of them is quite satisfactory: they share a defect that, rather interestingly, Tichý's own definition manages to avoid.

Peter Mott pointed out to me that distance from the truth, according to Tichý's measure $\triangle$, may change as we proceed from one language to a richer one. We have already seen an example of this: $\triangle_n(\mathit{taut}) = n/2$. This is perhaps not a very surprising result, and nor is it, on its own, very serious. For 'the truth' itself changes as we enrich our language. As 'the truth' grows in stature, it is natural that a mere tautology should slink further and further into obscurity. Moreover the absolute values of $\triangle$ are not in themselves of any great importance; what are important are the comparisons $\triangle$ yields of relative truthlikeness. And these, it transpires, are invariant.

For suppose that a is a sentence in a language with n atomic sentences, and that, as before, a commits k atomic errors in its m constituents. Then $\triangle_n(a) = k/m$. If a further atomic sentence is added to the language then the new disjunctive normal form of a will contain $2m$ constituents—for each old constituent must be conjoined separately with the new atomic sentence and its negation to provide the new constituents. Each of the k original atomic errors will therefore occur twice, and there will in addition be m further errors. Thus $\triangle_{n+1}(a) = (2k+m)/2m = \triangle_n(a)+1/2$. It is obvious that $k/m < k'/m'$ if and only if $(2k+m)/2m < (2k'+m')/2m'$. Thus comparisons of distance from the truth are unaffected by any extension of the language.

As a simple example, consider a language with two atomic sentences h and r. Let a be the sentence $\neg h \mathbin{\&} \neg r$ and b the sentence $(\neg h \mathbin{\&} r) \vee (h \mathbin{\&} \neg r)$. Then $\triangle_2(a) = 2 > 1 = \triangle_2(b)$. When a third atomic sentence w is included, a becomes $(\neg h \mathbin{\&} \neg r \mathbin{\&} w) \vee (\neg h \mathbin{\&} \neg r \mathbin{\&} \neg w)$, so that $\triangle_3(a) = 5/2$. The sentence b has disjunctive normal form $(\neg h \mathbin{\&} r \mathbin{\&} w) \vee (\neg h \mathbin{\&} r \mathbin{\&} \neg w) \vee (h \mathbin{\&} \neg r \mathbin{\&} w) \vee (h \mathbin{\&} \neg r \mathbin{\&} \neg w)$, so that $\triangle_3(b) = 6/4$. As desired, $\triangle_3(b) < \triangle_3(a)$.

However $\triangle'$ and $\triangle''$ do not share this desirable property. If $\triangle'_n(a) = k/(m+2^{n-1}(n-2))$ then $\triangle'_{n+1}(a) = (2k+m)/(2m+2^n(n-1))$. In the example above $\triangle'_2(a) = 2 > 1 = \triangle'_2(b)$. Again $\triangle'_3(a) = 5/6 > 6/8 = \triangle'_3(b)$. But if we now add a fourth atomic sentence we obtain

$$\triangle'_4(a) = (10+2)/(4+16) = 3/5 < 2/3 = (12+4)/(8+16) = \triangle'_4(b).$$

Thus the order of closeness to the truth has been reversed. A similar result clearly holds for $\triangle''$.

Tichý's arguments against translatability recorded above would no doubt allow him to make light of this lack of constancy. Niiniluoto too has recently claimed ([1976], section 3) that judgments of relative truthlikeness need not be persistent under language extension; it is doubtful, however, whether his arguments, even if valid, would be applicable in this case. One can of course argue that as 'the truth' is different in different languages it is far from impossible that one theory should be better at encapsulating the truth in one language, whilst the other theory is better for the other language. But though this is unexceptionable as a possibility, it would need some argument to show that all reversals of the sort noted above are defensible on this basis. My own view, therefore, is that, even if it is not as devastating as the objection Popper and Mott make to $\triangle$, this is a serious criticism of the measures $\triangle'$ and $\triangle''$. Whether there is any measure, akin to Tichý's, that suffers neither disadvantage,[1] I do not know. But even if there is, it will surely be as open to the complaint I originally raised as all the others are.

---

[1] The advantage of $\triangle''$ over $\triangle'$ is that it satisfies a special case of the *strict monotony condition*: if a and b are true and $b \vdash a$ then b is definitely closer to the truth than a is. This does not hold for $\triangle'$ if a is *taut* and b is obtained from it by dropping a disjunct that contains a single mistake.

3 On the final page of his [1974] Tichý hinted that the theory adumbrated above for sentential languages could be extended to general elementary languages by replacing the disjunctive normal form of a sentence with what Hintikka ([1953]) calls its ***distributive normal form.*** Suppose that we have a language without constants and with only finitely many primitive ***singulary*** predicates $P_0, \ldots, P_{n-1}$. From these can be generated 2^n Q-predicates (Carnap [1950], pp. 122–6) in the usual way.[2] A ***constituent*** is a sentence that asserts for each Q-predicate whether or not it is instantiated. Constituents are, as before, the maximal consistent sentences of the language. Hintikka showed that ***every consistent sentence is expressible as a disjunction of constituents***, its distributive normal form. (Basically this is a special case of THEOREM 36 of Tarski [1936].) For languages with binary predicates and beyond, the analysis becomes immensely more involved, largely because there is no limit to the number of definable 'Q-predicates'; this is obvious when one considers the number of relations definable in the language of Zermelo/Fraenkel set theory, which contains a single binary predicate. Hintikka therefore proposes that we call ***constituents at depth*** d the strongest consistent sentences formulable with at most d layers of quantifiers. It is then possible to cast each sentence with at most d layers of quantifiers into ***distributive normal form at depth*** d; that is, to write it as a disjunction of constituents at depth d.

When all the predicates are singulary there is one particularly natural way of extending Tichý's original treatment. As θ runs through the Q-predicates the 2^n sentences $\exists x \theta x$ are completely independent,[3] and each constituent can be expressed as some conjunction of them together with their negations. In effect, therefore, we are considering a sentential language with 2^n atomic sentences. Tichý's measure $\triangle$ can be defined in the same way as before.

Tichý does not take this path. (It has recently been explored by Niiniluoto [1976].) Instead he adopts a course which works as well for binary (and other more-than-singulary) predicates as it does for singulary ones. It cannot be pretended, however, that Tichý's definitions, presented in his section 3, are especially simple to understand, despite the examples of section 1 and the 'intuitive background' of section 2. I shall try therefore to give a flavour of the adamantine complexity of Tichý's work by spelling out the details of one of the most rudimentary examples possible. As might have been expected, I shall use the example to draw conclusions distinctly unfavourable to the whole enterprise.

---

[1] Professor Niiniluoto has pointed out to me that two of the measures of his [1976] are of this sort if restricted to the sentential case.

[2] For each primitive predicate P_i a Q-predicate ascribes either P_i or $\neg P_i$ to its argument. For example, $\theta x = P_0 x \mathbin{\&} \ldots \mathbin{\&} P_{n-1} x$ is a Q-predicate. Note that, in contrast to the sentential case, the finitude of the language is here essential.

[3] (*Added in proof.*) This is not quite true. The 2^n sentences cannot all fail simultaneously.

The example is Tichý's EXAMPLE B. Here we have a language with two primitive singulary predicates P and Q, as well as the sign $=$ of identity. The presence of identity allows us to assert the cardinalities of finite sets; in particular it allows us to assert, for any k, that there are at least k distinct elements in the universe.

Tichý supposes that 'in actual fact P and Q have two-element extensions which partially but not entirely overlap' (p. 26). He supposes further that the 'universe of discourse is a denumerably infinite collection' (*loc. cit.*). Since every sentence of the form 'there exist at least k distinct elements' is therefore true, it follows that 'the whole truth' is not expressible by any single sentence of this language.

Indeed, any complete finitely axiomatisable theory expressed in this language will ascribe a finite character to the whole universe. Tichý seems to suppose that this should be ruled out *a priori*; his whole 'sampling method'[1] described in his section 2 relies to some extent on the universe's not being too small. What we can do, of course, is to consider only denumerable interpretations. In this event, a theory like ($B3$) below, which says that P is satisfied by exactly two objects and Q by exactly one, and that these objects are distinct, can be regarded as complete; for no consistent strengthening of it rules out any further interpretations. In discussions of EXAMPLE B *I shall adopt this convention for the remainder of the paper.*

Tichý considers four sentences of the language and their relative distance from the truth ($\exists_n x A$ is an abbreviation 'for the standard formula saying that exactly n values of x satisfy A'[2]):

$$(B1)\quad \exists_1 x\, Px \,\&\, \forall x \neg Qx$$

$$(B2)\quad \exists_2 x\, Px \,\&\, \forall x \neg Qx$$

$$(B3)\quad \exists_2 x\, Px \,\&\, \exists_1 x\, Qx \,\&\, \forall x\, (Px \rightarrow \neg Qx)$$

$$(B4)\quad \exists_2 x\, Px \,\&\, \exists_2 x\, Qx \,\&\, \forall x\, (Px \rightarrow \neg Qx).$$

($B3$) and ($B4$) are the only members of this quartet with which I shall concern myself.

It is at this point perhaps helpful to interpose a simple diagram that shows what it is that ($B3$) and ($B4$) assert, and what is the truth (T) of the matter.

---

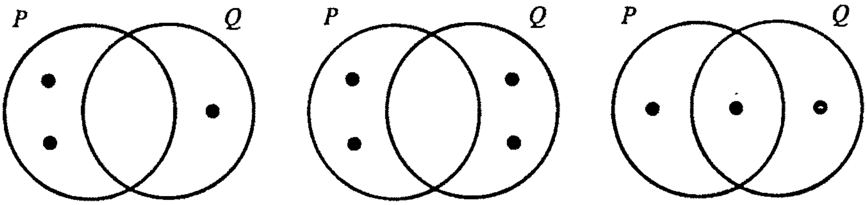
[1] I suppose that Tichý will now accuse me of speaking 'disparagingly', and will impute to me the 'implication . . . that it is not proper for a logician to stoop to [sampling] . . . things' (see his p. 33). I can only say that I do not normally interest myself in the propriety, or lack of it, of logicians' activities—only their validity.

Incidentally, Tichý's view that 'one either counts . . . or one does not arrive at a *number*' (*loc. cit.*) is simply false. It is refuted in the clearest possible way by measure theory. As is well known, the measure of a set is in no unambiguous way related to its cardinality. For instance, a set of the power of the continuum may have any measure from 0 (Cantor's discontinuum) to 1 (its complement). *Counting* the elements of an uncountable set—even supposing that such a thing could be done—provides no clue to its measure.

[2] I have here and elsewhere made some inconsequential notational changes when quoting from Tichý's paper, where, as it happens, the string '. Px' has been omitted from the statement of ($B3$). This has been done to make the business of comparison a little less messy.

It will become clear presently why I have drawn overlapping circles even when there are no elements in the common ground.

The strategy of Tichý's paper is: first to present examples that 'elicit . . . [the] intuition underlying the concept of verisimilitude' (p. 25); then to provide the details of his definition of verisimilitude (sections 2 and 3); and then to check that the measures afforded by his definition are 'in full accord with what is intuitively expected' (p. 36). In my criticisms I shall follow the same order. I shall first argue that Tichý's intuitive judgments with regard to EXAMPLE B are untrustworthy and based on an imperfect appreciation of the complexities involved. I shall then reformulate his definition in a way that (I hope) is a bit more easily understood than his own presentation. Finally, I shall show that even if Tichý's intuitive assessments are perfectly correct, his definition cannot be relied on to reflect them.



($B3$) ($B4$) (T)

Figure 1

After introducing the theories ($B1$)–($B4$) Tichý considers how they stand in relation to the truth (T). 'As for ($B4$),', he writes (p. 26), 'it corrects something which ($B3$) gets wrong, *viz.* the size of Q; on the other hand, it gets wrong something which ($B3$) gets right, *viz.* the number of Q's which are outside P. ($B3$) and ($B4$) thus seem equidistant from the truth'. In other words, Tichý sees ($B3$) and ($B4$) as engaged in answering a set of three questions: 'What is the cardinality of P?', 'What is the cardinality of Q?', and 'What is the cardinality of $Q\ \&\ \neg P$?'. Using sentential connectives as part of an obvious shorthand, we can represent their answers, and the truth of the matter, thus:

($B3$a)	($B4$a)	(Ta)
$P = 2$	$P = 2$	$P = 2$
$Q = 1$	$Q = 2$	$Q = 2$
$Q\ \&\ \neg P = 1$	$Q\ \&\ \neg P = 2$	$Q\ \&\ \neg P = 1$

It is clear that ($B3$) and ($B4$) both offer two correct answers and one wrong one to these three questions. In agreement with the principles underlying his sentential measure, Tichý accordingly judges ($B3$) and ($B4$) equally distant from the truth.

I surely cannot be the only person who finds this reasoning somewhat less than compelling. For what it is worth—and I shall argue that it is not worth much—my 'intuition' tells me that the three questions ($B3$) and ($B4$) should be taken as responding to are 'What is the cardinality of P?', 'What is the cardinality of Q?', and 'What is the cardinality of $P \& Q$?'. The responses can be tabulated in the same way:

($B3$b)	($B4$b)	(Tb)
$P = 2$	$P = 2$	$P = 2$
$Q = 1$	$Q = 2$	$Q = 2$
$P \& Q = 0$	$P \& Q = 0$	$P \& Q = 1$

So looked at, ($B4$) seems by Tichý's principles to be distinctly better than is ($B3$).

There are several other ways of viewing ($B3$) and ($B4$) as answers to three different cardinality questions. In the following table ($B3$) is more often correct than is ($B4$):

($B3$c)	($B4$c)	(Tc)
$P = 2$	$P = 2$	$P = 2$
$P \vee Q = 3$	$P \vee Q = 4$	$P \vee Q = 3$
$P \& \neg Q = 2$	$P \& \neg Q = 2$	$P \& \neg Q = 1$

Thus depending on what questions we take as fundamental we can arrive at any one of three possible judgments: (a) ($B3$) and ($B4$) are equally close to the truth; (b) ($B4$) is closer to the truth than is ($B3$); (c) ($B3$) is closer to the truth than is ($B4$).

The reader will have no difficulty in checking that the three sets of questions are equivalent, in the sense that answers to one set of questions automatically yield answers to the other two sets. Moreover, none of the sets contains redundant elements. From a knowledge of the cardinalities of P and Q, for instance, it is not possible in general to calculate how many elements satisfy $Q \& \neg P$, or $P \& Q$, or $P \& \neg Q$. (There are of course special cases: if $Q = 0$ then both $Q \& \neg P$ and $P \& Q$ must be zero.) Can there be any reason for preferring one of these sets of questions to the other two?

There can be no recourse here, as there was in the sentential case, to an appeal to *linguistic primitiveness*. P and Q are primitive predicates, but none of $Q \& \neg P$, $P \& Q$, $P \vee Q$, $P \& \neg Q$ is. There is nothing in the language we employ that could indicate that $Q \& \neg P$, rather than $P \& Q$, is the third predicate whose cardinality ($B3$) and ($B4$) are trying to provide. The fact is that there is a crucial difference between sentential languages and those occupying our attention at present. *A complete theory* (*constituent*) *in a sentential language is a conjunction of the elements of a completely independent set of sentences*, namely atomic sentences and their negations. Every conjunct involves a different primitive. In predicate languages however, a complete theory such as ($B3$) or ($B4$) (see p. 369 *above*) is not a conjunction of sentences each of which involves a different primitive predicate. To be sure, such a theory *is* a conjunction of completely independent sentences. One can even say, though I have rather pointedly tried to avoid saying it until now, that ($B3$) or ($B4$) is a conjunction of answers to *a set of independent cardinality questions*. But these 'independent' cardinality questions do not ask for the cardinalities of independent predicates, A theory which merely told us the cardinalities of P and Q would simply not tell us enough.

There is a problem here that has never been satisfactorily posed, as far as I know, let alone solved: ***under what conditions is a set of predicates independent?*** This way of posing the problem is not satisfactory, because it suggests that we are concerned with 'explication', that is, with a basically verbal puzzle. It is my view, however, that there is a substantive question here, and that answering it yields some unexpected dividends. (For some comments on 'explication' see the last section of this paper.)

This is not the place to discuss how the problem might be solved. (I hope to say something about it elsewhere.) For present purposes it is sufficient to quote the solution (or what I take to be the solution) for two singulary predicates P, Q: P and Q are independent if any combination whatever of cardinalities (apart from four zeros) is possible for the four Q-predicates $P \,\&\, \neg Q$, $P \,\&\, \neg Q$, $Q \,\&\, \neg P$, $\neg Q \,\&\, \neg P$. More formally: if λ, μ, ν, ξ, are numerals, not all numerals for zero, then the four sentences 'The cardinality of $P \,\&\, Q$ is λ', 'The cardinality of $P \,\&\, \neg Q$ is μ', 'The cardinality of $Q \,\&\, \neg P$ is ν', 'The cardinality of $\neg Q \,\&\, \neg P$ is ξ' are mutually consistent. If this solution is correct, it seems to follow that in Tichý's EXAMPLE B the *fundamental* cardinality questions concern the four Q-predicates definable from the primitives P and Q. Now as it happens Tichý's restriction to denumerable domains technically speaking upsets the independence of P and Q, according to what I say; for ***if the domain is denumerable, at least one of the Q-predicates must be denumerable.*** But this is not really a relevant consideration since that part of the domain outside both P and Q plays no role in the example anyway. Thus, I would argue, there are in this example three fundamental cardinality questions: 'What is the cardinality of $P \,\&\, \neg Q$?', 'What is the cardinality of $P \,\&\, Q$?', and 'What is the cardinality of $Q \,\&\, \neg P$?'.[1]

A new table shows how pitifully ($B3$) and ($B4$) perform over *these* questions:

($B3$d)	($B4$d)	(Td)
$P \,\&\, \neg Q = 2$	$P \,\&\, \neg Q = 2$	$P \,\&\, \neg Q = 1$
$P \,\&\, Q = 0$	$P \,\&\, Q = 0$	$P \,\&\, Q = 1$
$Q \,\&\, \neg P = 1$	$Q \,\&\, \neg P = 2$	$Q \,\&\, \neg P = 1$

The reader may consult Figure 1, reading from left to right; he will understand now why the circles were drawn overlapping in every case. ($B4$), it can be seen, gets everything wrong, whilst ($B3$) gets one answer out of three right. Theories that would score higher marks if we simply counted correct answers include[2]

$$(B5) \quad \exists_1 x\, Px \,\&\, \exists_1 x\, Qx \,\&\, \forall x\, (Px \rightarrow \neg Qx)$$

$$(B6) \quad \exists_2 x\, Px \,\&\, \exists_1 x\, Qx \,\&\, \forall x\, (Qx \rightarrow Px)$$

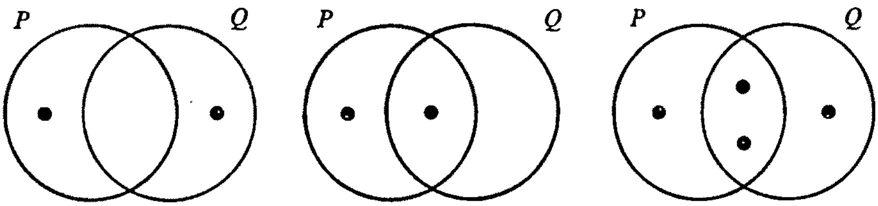
$$(B7) \quad \exists_3 x\, Px \,\&\, \exists_3 x\, Qx \,\&\, \exists_2 x\, (Px \,\&\, Qx),$$

---

[1] (*Added in proof.*) Professor Watkins has suggested that we consider not just a *fundamental* set of cardinality questions, but *all* cardinality questions. It is not hard to see that not much is changed by this tactic (with which I am rather sympathetic).

[2] ($B5$) has been so named by Tichý on p. 39 of his [1976]. ($B6$) and ($B7$), and ($B8$) and ($B9$) below, are my own inventions.

which can be represented diagrammatically in a similar way:



($B5$) ($B6$) ($B7$)

Figure 2

The fundamental questions are answered as follows:

($B5$)	($B6$)	($B7$)
$P \,\&\, \neg Q = 1$	$P \,\&\, \neg Q = 1$	$P \,\&\, \neg Q = 1$
$P \,\&\, Q = 0$	$P \,\&\, Q = 1$	$P \,\&\, Q = 2$
$Q \,\&\, \neg P = 1$	$Q \,\&\, \neg P = 0$	$Q \,\&\, \neg P = 1$

Let me sum up this part of my argument. Although the atomic sentences in a sentential language in no way single out a unique set of *independent* questions that the complete theories of the language can be taken as answering, something like this is true of cardinality questions in a predicate language with equality. Whether the primitive predicates are P, Q or P, $P \leftrightarrow Q$, the same *fundamental* set of cardinality questions will arise. But these questions do *not* directly concern the cardinalities of the *primitive* predicates themselves. Tichý seems to assume that they do; and, as confessed above, that is what my 'intuition' told me too. Tichý presumably further thinks that there is a reason—though he makes no attempt to state it—for adopting $Q \,\&\, \neg P$ in addition to P and Q. Given that the adoption even of $P \,\&\, \neg Q$ would beget a different result (($B3$) would have two false answers and ($B4$) would have one), I find it unimaginable what this reason might be. (See also the remarks in the next paragraph.) Tichý's 'intuitive comparisons', therefore, seem not only utterly arbitrary, but even inconsistent. In any case, they are surely incorrect.

4 However, *they are only 'intuitive comparisons'*; *they do not constitute Tichý's theory of truthlikeness.* One would of course expect Tichý's theory to vindicate his intuitive judgments; and in effect he claims that they do when he reports, without any details of the calculation, that 'the distances of ($B1$), ($B2$), ($B3$), and ($B4$) from the actual world are 2.57, 2, 0.62, and 0.62, again in full accord with what is intuitively expected' (p. 36). Yet closely similar comparisons are not ratified by the theory. Of these, much the easiest to follow is the comparison of ($B6$) with

$$(B8) \quad \exists_1 x\, Px \,\&\, \exists_2 x\, Qx \,\&\, \forall x\, (Px \rightarrow Qx),$$

which is simply what one obtains from ($B6$) by interchanging 'P' and 'Q'. Since (T) is symmetrical in P and Q it is immediate that in any formal theory of verisimilitude—such as Tichý's—($B6$) and ($B8$) will be equally close to (T). Yet ($B6$) gets correct only the cardinality of P, whilst ($B8$) gets correct the cardinalities of both Q and $Q \,\&\, \neg P$. By Tichý's reckoning, therefore, ($B8$) should be closer to the truth than is ($B6$). On the other hand, if the fundamental cardinalities are, as I have suggested, those of $P \,\&\, \neg Q$, $P \,\&\, Q$, and $Q \,\&\, \neg P$, then ($B6$) and ($B8$) are equally close to the truth, since each scores two points out of three.

If my calculations are correct[1] there are a number of distinctly more impressive (and much less swiftly established) discrepancies between Tichý's theory and his intuition. In particular, we can show that

$$(B9) \quad \exists_2 x\, Px \,\&\, \exists_2 x\, Qx \,\&\, \forall x\, (Qx \rightarrow Px),$$

which correctly propounds the cardinalities of both P and Q, has distance 42/34 from the truth, compared with 21/34 for ($B3$), ($B4$) and ($B6$). This is especially serious for the comparison of ($B9$) and ($B6$); for it shows that a theory with two correct answers can be considerably more distant from the truth than a theory with only one. Indeed, since ($B5$) has distance 27/34 from the truth it is possible for a theory that gets correct the cardinalities of both the primitive predicates (P and Q) to be a lot further from the truth than one that gets neither correct.

Might it not be that these calculations merely demonstrate once more the unreliability of Tichý's intuitive comparisons, and speak neither for nor against the formal theory? I think not. For it can easily be shown that the theory does very little better when we consider, as I have suggested, the fundamental cardinalities to be those of $P \,\&\, \neg Q$, $P \,\&\, Q$, and $Q \,\&\, \neg P$, the three undivided areas of each diagram. The three sentences ($B5$), ($B6$), ($B7$), each of which gets one cardinality wrong (by only 1 in each case), have distances from the truth of 27/34, 21/34, 9/43 respectively. It is not clear in what way ($B7$) is so noticeably better than ($B5$) or ($B6$). Again, ($B6$), ($B3$), ($B4$) are all the same distance from the truth (according to the theory), even though they make respectively one error, two errors, three errors. No recognisable pattern of good behaviour seems to emerge from the intricacies of this benighted example.

There is one more difficulty that is worth mentioning before we get down to work. As the theory is formulated in Tichý's paper (see Definition 19 on p. 33) comparisons of greater or less verisimilitude between two sentences are based on their comparative distances Δ_d from the truth evaluated at depth d (the depth of the deeper of the two sentences). Now ($B5$) and ($B6$) have depth 2, whilst ($B3$) has depth 3.[2] The distances reported above were all calculated at depth 3. Thus $\Delta_3(B6) = 21/34 = \Delta_3(B3) < 27/34 = \Delta_3(B5)$, so that ($B3$) and ($B6$) are equal in verisimilitude and ($B3$) has more verisimilitude than ($B5$). It would seem reasonable to deduce that ($B6$) has more verisimilitude than ($B5$). But, on the contrary, $\Delta_2(B6) = \Delta_2(B5)$, so that ($B5$) and ($B6$) have equal verisimilitude. Now it may be that variations of this sort can be damped out by plunging to greater depths. For at depth 4 the distances of $(B3)$, $(B5)$ and $(B6)$ from the truth are 52/73, 76/73 and 52/73 respectively. But, as it stands, Tichý's theory is barely consistent. This alone shows that some alteration is urgently needed.

---

[1] Were I an inductivist I would have to regard this sentence as counterfactual (which perhaps it is). In trying to get to the heart of Tichý's theory I have attempted to repeat his computations, and have frequently ended up with the wrong answers. I should like to acknowledge my indebtedness to Professor Tichý for patiently correcting my blunders on a number of occasions.

[2] As formulated on p. 372 *above*, ($B6$) appears to have depth 3. But it may also be written $\exists_1 x\, (Px \,\&\, \neg\, Qx) \,\&\, \exists_1 x\, (Px \,\&\, Qx) \,\&\, \forall x\, (Qx \rightarrow Px)$, which is of depth 2. It is clearly the minimum depth that we must use for each sentence, for there is no maximum.

Tichý's measure of distance from the truth is limited to theories that are finitely axiomatisable; that is, to sentences. In my exposition I propose to confine myself further to *complete* sentences (in the sense specified on p.372 *above*), such as $(B3), \ldots, (B9)$. (The truth (T) is itself a complete theory, of course, and in EXAMPLE B it is also finitely axiomatisable. But this restriction is inessential.) If X and Y are complete sentences, and d is the depth of the deeper one (or of both if they are equally deep), Tichý asks us to consider three sets of ***ordered samples*** of size d: those ordered samples that are compatible with X (that is, could be drawn from a universe characterised by X), those compatible with Y, and those compatible with T. As a simple example, consider $(B5)$ and $(B6)$, each of which has depth 2.[1] We list all varieties of ordered pairs that are compatible with each theory, and with the truth. A glance at the diagram of $(B5)$ will show that if $\langle a, b\rangle$ is a pair drawn from this universe then one of seven possibilities obtains (each ordered sample occupies one row):

$P \,\&\, \neg Q$	$P \,\&\, Q$	$Q \,\&\, \neg P$	$\neg P \,\&\, \neg Q$
a		b	
b		a	
a			b
b			a
		a	b
		b	a
			a, b

I have bracketed together samples that are merely permutations of one another; they have the same distribution, and differ only in their order of presentation. Here we may read 'a' as an abbreviation of 'the first element sampled', and 'b' as an abbreviation of 'the second element sampled'. It is clear that the number of possible samples can be quite considerable (for $(B4)$, which is of depth 3, we have to consider 34 triples), and a convenient method of abbreviation is needed. Referring to the three undivided areas of each diagram I shall write *m–n–p* to indicate that there are m objects satisfying $P \,\&\, \neg Q$, n objects satisfying $P \,\&\, Q$, and p satisfying $Q \,\&\, \neg P$. I shall not distinguish different permutations, but will simply note on the right how many of each there are. The following table therefore catalogues all the ordered pairs that are compatible severally with $(B5)$, with (T), and with $(B6)$.

Each sample of size d can, in the case of a language with k primitive singulary predicates, be represented as a conjunction of $k.d$ completely independent sentences, namely atomic sentences or their negations. Thus the two samples of type 1–1–0 allowed by (T) are $Pa \,\&\, \neg Qa \,\&\, Pb \,\&\, Qb$ and $Pa \,\&\, Qa \,\&\, Pb \neg Qb$. Given Tichý's measure of distance between sentential constituents (criticised in 1 above) we should not be surprised at the way he measures the distance between two samples of size d: it is simply the number of atomic sentences about which the two samples differ. Thus the two samples named above have distance 1 each from the samples $Pa \,\&\, \neg Qa \,\&\, \neg Pb \,\&\, Qb$ and $\neg Pa \,\&\, Qa \,\&\, Pb \,\&\, \neg Qb$ respectively, each of which is of type 1–0–1 and permitted by ($B5$). (Note that if two samples are of the same type, that is, obtained from each other by permutation of individual constants, and one of them is distant δ from a member of some other type, then the other is also distant δ from a member of that other type.) Indeed, it is easily checked that every sample in (T) has distance of at most 1 from some sample in ($B5$).

---

[1] See the preceding footnote.

($B5$)	(T)	($B6$)	Number
	1–1–0	1–1–0	2
1–0–1	1–0–1		2
	0–1–1		2
1–0–0	1–0–0	1–0–0	2
	0–1–0	0–1–0	2
0–0–1	0–0–1		2
0–0–0	0–0–0	0–0–0	1

An undemanding method for computing the distance between two samples is obtained by reading the notation m–n–p for a type of sample as an assertion of the number of objects in each of three adjacent boxes. On the outer side of each of the two outer boxes there is an unexplored background of objects that satisfy neither P nor Q. (Alternatively, we can think of four boxes arranged in a ring.) Call any transfer of an object from one box to an adjacent one, or from an endbox into the background, or from the background into an endbox, a ***move***. Then the distance between two samples is 1 if and only if a single move takes one sample into the other. More generally, the distance between two samples is the smallest number of moves needed to get from one to the other. 1–1–0 and 0–1–1 are distance 1 from 1–0–1; 0–1–0 is distance 1 from 1–0–0; and all the other samples in (T) have zero distance from themselves in ($B5$).

Tichý defines a ***linkage*** between two sets or columns of samples as a relation whose domain is one column, whose range is the other, and which is ***minimal*** in the sense that no relation of lower cardinality has the same domain and range ([1976], Definition 3, p. 30). A linkage therefore is simply a surjective mapping from the larger column to the smaller. A moment's reflection shows that a linkage always has just as many links as there are samples in the larger column. Since (T) has 13 pairs (including permutations), and ($B5$) only 7, a linkage between the two columns has 13 links. The same is true for (T) and ($B6$). The ***breadth*** of a linkage is the average distance between the pairs that constitute it (Definition 2); in other words, the average length of link.

Tichý defines the distance between two sets of samples, and in effect the distance between the two theories, as ***the breadth of the narrowest linkage between them*** (Definition 4). The distance between T and ($B5$) is therefore 6/13, since seven links have length 0 and six have length 1. Likewise the distance between (T) and ($B6$) is 6/13. For 1–0–1 has distance 1 from 1–1–0; and each of 0–1–1 and 0–0–1 have distance 1 from 0–1–0. Again there are six unit links and seven zero ones. Thus at depth 2 the sentences ($B5$) and ($B6$) are equally close to the truth.

Now consider the comparisons of (B3), (B4), (B5), (B6), (B7), and (B9) at depth 3. The sample types can be read very easily off the diagrams on pages 370 and 373 *above*. (In the case of (B9) a diagram is hardly needed.)

(B3)	(B4)	(B5)	(B6)	(B7)	(B9)	(T)	Number
2–0–1	2–0–1						3
2–0–0	2–0–0						3
				1–2–0			3
				1–1–1		1–1–1	6
			1–1–0	1–1–0		1–1–0	6
	1–0–2						3
1–0–1	1–0–1	1–0–1		1–0–1		1–0–1	6
1–0–0	1–0–0	1–0–0	1–0–0	1–0–0		1–0–0	3
				0–2–1			3
				0–2–0	0–2–0		3
				0–1–1		0–1–1	6
			0–1–0	0–1–0	0–1–0	0–1–0	3
	0–0–2						3
0–0–1	0–0–1	0–0–1		0–0–1		0–0–1	3
0–0–0	0–0–0	0–0–0	0–0–0	0–0–0	0–0–0	0–0–0	1

(T) permits a greater number of distinct samples—namely 34—than any theory but (B7). Thus the linkages between (T) and any other theory—save (B7)—each contain 34 links.

Consider first (T) and (B4). Three of the six 1–1–1 samples in (T) can be linked with the three 2–0–1 samples (unit distance), whilst the other three are linked with the three 1–0–2 samples (unit distance). The six 1–0–1 samples in (T) can be linked in pairs to the three 2–0–0 samples in (B4) (unit links again). Likewise the six 0–1–1 samples in (T) are unit distance from the three 0–0–2 samples in (B4). The three samples of type 0–1–0 are distance 1 from the three samples of type 0–0–1. All the other samples in (T) are zero distance from themselves in (B4). By this method every sample in each of the two lists is linked, by a link of length at most unity, to one in the other list. Since there are 34 links, of which 13 have length 0, the distance between (B4) and (T) is 21/34. This agrees with Tichý's value of 0.62.

The reader will find it easy to confirm that, in the same way, every sample in (T) is at most unit distance from one in (B3). The converse needs no checking, since every (B3) sample is a (B4) sample. By an identical argument we can conclude that the distance between (B3) and (T) is also 21/34. Exactly similar considerations apply in the case of (B6).

With (B5), however, there is some slight novelty. The six samples of type 1–1–1 in (T) do not have unit distance from any samples in (B5). They are, on the other hand, of distance 2 from the 1–0–1 samples. This adds 6 to the total length of links involved in the narrowest linkage between (T) and (B5). Thus (B5) is distant 27/34 from (T).

There are 43 triples in ($B7$), and the 9 of them not compatible with (T) are obviously unit distance from something in (T). Thus ($B7$) has distance 9/43 from the truth. A glance at the table, however, should convince one that the distance between ($B9$) and (T) is very much larger. The reader may check that it is in fact 42/34. I also leave to the reader the task of checking the distances at depth 4 of ($B3$), ($B5$), and ($B6$) from (T); I would note only that no new sample *types* are here introduced by going to the greater depth—all we do is increase the number of samples in most types. This indicates that one *prima facie* attractive way of resolving the difficulty about ($B5$) and ($B6$) mentioned on pages 374 f. above is likely to be unsuccessful in general. This is the suggestion that the distance between two samples should be measured by the number of Q-predicates they disagree on, not the number of primitives. Such a measure would also have the advantage, in contrast to Tichý's own, of not being vulnerable to the all-too-obvious charge of language-dependence.

But perhaps Tichý would pay no attention to this suggestion anyway. For it so seriously detracts from the generality of his formal theory as to make that theory virtually unnecessary. Rather than embroil ourselves in the unfathomable complexities of sampling, we might just as well return to, and take our stand on, the 'intuitive comparisons' that the theory set out to authorise; if not those of Tichý's section 1, then the corrected ones I offered above. Doubtless these comparisons are not all that there is to verisimilitude. But they are something. And they are intelligible.

5 By suggesting that Tichý's or my 'intuitive comparisons' can be regarded by themselves as a primitive theory of verisimilitude I do not of course wish to imply that these comparisons are correct, let alone that they are exempt from critical scrutiny. Indeed, the opposite is the case: treating them as theoretical serves to emphasise how conjectural, and criticisable, these 'intuitive comparisons' really are. I would have thought that a wariness about 'intuitions', whether or not 'remarkably distinct' (Tichý [1976], p. 25), would have been universal in these days. Yet throughout his paper [1976] Tichý writes as though 'the intuition underlying the concept of verisimilitude' were an unimpeachable one, and that the comparisons offered in his three examples were incontestably correct. (I intend to say nothing about EXAMPLE C, since I have discussed similar examples at length in my [1975*a*] and [1975*b*].) For instance, regarding his EXAMPLE A (the *h–r–w* example outlined on pp. 364 f. *above*) he says (p. 33): 'It seems to me that if ($A1$) [that is, $\neg h \,\&\, \neg r \,\&\, \neg w$] and ($A4$) [that is, $\neg h \,\&\, r \,\&\, w$] do not constitute an instance of one theory['s] being closer to the truth than another [is], then nothing does'. He then records (p. 34) that I had already presented an argument against this judgment, and quotes me rather scandalously out of context by writing that my argument 'if generalised, purports to show that it is impossible to "make . . . correct judgments of . . . comparative proximity to the truth" at all'.[1] Actually the argument neither purported to show, nor did show, anything of the sort. It purported to show,

---

[1] The passage in double quotes here, 'make . . . correct judgments of . . . comparative proximity to the truth' appears in full in my [1974*a*] on p. 175: 'Both Tichý and Harris claim to be able to make, in certain very simple situations, correct judgments of the comparative proximity to the truth of different theories.'

and indeed did show, that Tichý's own intuitive judgments were seriously defective.

Tichý proceeds (*loc. cit.*) to investigate the mystery of how I could possibly maintain what I do actually maintain:

> One way to construe Miller's position is as a confession that he, Miller, simply lacks my intuition that ($A4$) is closer to the truth than ($A1$) is .... If this construal was correct, ... it would be quite inexplicable why he should consider verisimilitude a worthwhile topic in the first place; trying to explicate a putative notion which lacks a single clearcut intuitive instance ... is hardly a serious venture. Finally, it would be totally incomprehensible why Miller should find Popper's 'qualitative' definition of verisimilitude 'quite untenable'.... For implying, as it does, that no false theory is closer to the truth than another theory, the definition would rather be in perfect agreement with what Miller intuitively expected.
>
> But if, on the other hand, Miller shares Popper's and my pretheoretical intuitions about verisimilitude—intuitions, by the way, which I have so far found in the intellectual armoury of every person I have spoken to—then he has misunderstood the import of his own ... argument....

This passage is so misleading that I feel obliged to take a little space to respond. Let me say first that of course I *had* the 'intuition' that $\neg h \,\&\, r \,\&\, w$ is closer to $h \,\&\, r \,\&\, w$ than $\neg h \,\&\, \neg r \,\&\, \neg w$ is. I am not a bit surprised that Tichý finds this idea endemic to the intellectual population. After all, countless competitions, examinations, and tests, quizzes and questionnaires, from the most weighty to the 'Are You Really Any Good In Bed?' variety, base their assessments in some way on the number of correct answers given. (For an independent published statement of the same idea see Haack [1974], pp. 62 f.) But I now regard this 'intuition' as simply refuted by the *h–m–a* transformation exhibited in my [1974*a*] (and, more generally, by THEOREM 5 of my [1974*b*]). It is interesting that Tichý does not say whether he acquainted the persons he spoke to with this counterargument of mine.[1]

However, unless we take quite literally Tichý's remark that 'if ($A1$) and ($A4$) do not constitute an instance of one theory['s] being closer to the truth than another [is], then nothing does', I do not see why, *had* I lacked Tichý's intuition, this would have meant that for me verisimilitude was without 'a single clearcut intuitive instance'. At the very beginning of my [1974*a*], in explaining the problem of verisimilitude, I noted that 'theories of motion ... starting with Plato and Aristotle and developing through Buridan and Oresme, Galileo and Descartes, to Newton and Einstein, do give the appearance (deceptive perhaps) of progress—of progress towards the truth' (p. 166). Later, in my [1975*b*], I unashamedly endorsed the view that 'some oversimplifications are very much better than others' (p. 207). A third, pretty clearcut, intuitive instance (which, of course, Tichý could not know that I would venture) is to be found on p. 367 *above*, where I say that I consider both $\triangle'$ and $\triangle''$ as better than Tichý's measure $\triangle$, even though 'they share a defect that... Tichý's own definition manages to avoid'. In short, I have no doubt that there are clearcut intuitive instances of greater and less truthlikeness. But I do not know whether any of these instances, though clear, are correct.

---

[1] The implications of this argument for education, and for many social science survey methods, are perhaps worth pondering. Indeed, it was from this direction that Batchelder and Narens [1975] independently came upon essentially the same result. See especially section 2 of their report.

This pinpoints, I think, a crucial methodological distinction between Tichý's approach to the problem and my own. He thinks that there is a notion that we are 'trying to explicate'. For us to have a proper idea of what we are trying to do, and whether we have succeeded, we need 'clearcut intuitive instances' of the concept. These instances must be effectively beyond dispute, for were it possible for us to challenge, or overturn, some or all of them, the explication would be 'hardly a serious venture'. In my own formulations of the problem, on the other hand, I have tried hard to stress that the problem is not just one of 'rigorous definition' of words or concepts. In my [1974*a*], p. 166, I posed the problem as 'What can there be about one false theory that makes it closer to the truth than is another?', a formulation which makes little contact with worries about rigour or precision. Even this, however, I regard now as an insufficiently critical formulation, presupposing (as it seems to) that progress towards the truth *is* possible. I would prefer now the neutral version of my [1974*b*], p. 178: 'Are some theories *closer to the truth* than others are; and if so, what is the objective determinant of their greater truthlikeness?' To this question one of the possible answers is 'No false theories are closer to the truth than others are'. I don't say that I would welcome this answer, and I do not in fact think it is true. Nevertheless it is surely admissible. Yet I do not see how it could be admissible if the problem is treated as one of 'explication'.

Against this it may be contended that to assess the adequacy of a theory of verisimilitude we are going to require intuitively obvious cases where one theory is more truthlike than is another. Thus, it will be said, I am needlessly exaggerating the difference between Tichý's method and my own. Now I shall not deny that such cases may be necessary. But *they must come after the theory, as a challenge, not before it as a constraint.* 'Explication', anyway as Tichý regards it here, is a metaphysical ally of inductivism, proceeding from indubitable raw 'data' to explanatory theories. I suggest as a better method, less likely to lead us into dogmatism, the metaphysical twin of falsificationism: this means proposing explanatory hypotheses about a domain (in our case, scientific progress), and thereafter seeing how well they account for the phenomena of the domain. Only in this way are we likely to make an imaginative advance towards solving what is clearly a most intractable problem.

In conclusion I would say only that in my [1977*a*] I offer a theory of distance from the truth that, weak as it is, does genuinely seem to be the sort of theory we are looking for. (At least, one can say that it provides the core of a satisfactory theory. Some of the results have been advertised already in Popper [1976], p. 152.) The very existence of this theory makes it plain that Tichý's linguistically bound approach to the problem is not as inevitable as he himself suggests.

DAVID MILLER
University of Warwick

REFERENCES

Batchelder, W. H. and Narens, L. [1975]: 'A Critical Examination of the Analysis of Dichotomous Data', *Social Science Working Paper*, 49, School of Social Sciences, University of California, Irvine.

Black, M. [1964]: *A Companion to Wittgenstein's 'Tractatus'*. Cambridge University Press.

Carnap, R. [1950]: *Logical Foundations of Probability*. University of Chicago Press.

Haack, S. W. [1974]: *Deviant Logic*. Cambridge University Press.

Hanson, N. R. [1961]: 'Are Wave Mechanics and Matrix Mechanics Equivalent Theories?', in H. Feigl and G. Maxwell (*eds.*): *Current Issues in the Philosophy of Science*. New York: Holt, Rinehart and Winston, pp. 401–25.

Hintikka, K. J. J. [1953]: 'Distributive Normal Forms in the Calculus of Predicates', *Acta Philosophica Fennica*, **6**.

Kanger, S. [1968]: 'Equivalent Theories', *Theoria*, **34**, pp. 1–6.

Miller, D. W. [1974*a*]: 'Popper's Qualitative Theory of Verisimilitude', *British Journal for the Philosophy of Science*, **25**, pp. 166–77.

Miller, D. W. [1974*b*]: 'On the Comparison of False Theories by their Bases', *British Journal for the Philosophy of Science*, **25**, pp. 178–88.

Miller, D. W. [1975*a*]: 'The Accuracy of Predictions', *Synthese*, **30**, pp. 159–91.

Miller, D. W. [1975*b*]: 'The Accuracy of Predictions: A Reply', *Synthese*, **30**, pp. 207–19.

Miller, D. W. [1977*a*]: 'On Distance from the Truth as a True Distance', forthcoming in *Acta Philosophica Fennica*.

Miller, D. W. [1977*b*]: 'New Axioms for Boolean Geometry', forthcoming in *Bulletin of the Section of Logic*. Institute of Philosophy and Sociology, Polish Academy of Sciences, Wrocław.

Miller, D. W. [1977*c*]: 'The Uniqueness of Atomic Facts in Wittgenstein's *Tractatus*', forthcoming in *Theoria*.

Niiniluoto, I. [1976]: 'On the Truthlikeness of Generalizations', *forthcoming*.

Popper, K. R. [1934]: *Logik der Forschung*. Wien: Springer.

Popper, K. R. [1967]: 'Quantum Mechanics Without "The Observer"', in M. Bunge (*ed.*): *Quantum Theory and Reality*. New York, Berlin, Heidelberg: Springer-Verlag, pp. 7–44.

Popper, K. R. [1976]: 'A Note on Verisimilitude', *British Journal for the Philosophy of Science*, **27**, pp. 147–59.

Tarski, A. [1936]: 'Grundzüge des Systemenkalkül, Zweiter Teil', *Fundamenta Mathematicae*, **26**, pp. 283–301. References are to the translation by J. H. Woodger in Tarski [1956].

Tarski, A. [1956]: *Logic, Semantics, Metamathematics*. Oxford University Press.

Tichý, P. [1974]: 'On Popper's Definitions of Verisimilitude', *British Journal for the Philosophy of Science*, **25**, pp. 155–60.

Tichý, P. [1976]: 'Verisimilitude Redefined', *British Journal for the Philosophy of Science*, **27**, pp. 25–42.

Wittgenstein, L. [1921]: *Tractatus Logico-Philosophicus*. London: Routledge and Kegan Paul.

LOGIC, LITERACY, AND PROFESSOR GELLNER

It is always a pleasant surprise for an author to find a critic who understands his philosophy, agrees with it, and shows some ability to develop it further. It is even more gratifying to meet a thinker who shares with him not just ideas but some other idiosyncracy, especially when the idiosyncracy is unpopular, and frowned upon by the profession. For years Lakatos and I were alone in our attempt to inject a little life, some personal note into philosophical debate. Since Imre's death nobody was left to support me in this enterprise. Now the review of my book in this *Journal*[1] introduces a writer who is not only willing to leave the narrow path of academic prose and reasoning but who has great talent in this direction, who is a master in the art of invective and who has considerably augmented the inventory of rhetorical techniques. Perhaps I should have been grateful for the support which my efforts seemed to receive from such unexpected quarters and should not have probed further, but unfortunately my pedantry got the better of my gratitude. I soon discovered that though the reviewer writes well he does not at all write correctly. His ability to give colour to his own ideas and impressions is marred by a striking blindness for the ideas, the motives, the procedures of others. His interpretations of my text are but rarely conscious distortions, as they would be with a sophisticated rhetorician, they are mostly simple errors of reading and comprehension. Indeed, I discovered that we have here not a well planned extension of the art of *rhetorical* argument, but side effects of an abortive attempt at a *rational* criticism. This being the case I unfortunately cannot praise Gellner for his rhetorical acumen but am reduced to the unpleasant task of enumerating trivial errors and misunderstandings. In the remarks that follow I shall try to sweeten this task as well as I can, both for myself, and for my readers. I shall concentrate on points that not only reveal Gellner's procedure but are also of general interest and whose discussion, one hopes, will achieve a little more than a mere return to the text from which Gellner started.

---

[1] Gellner [1975].

Gellner's review contains (*i*) a presentation of my main theses and arguments; (*ii*) a criticism of my style and an evaluation of its results; (*iii*) a sociological analysis of the 'happening' (p. 332) that is my book. I shall now discuss these points in order.

(*i*) At first sight it would seem that Gellner gives a fairly accurate presentation of what I say for the sentences he writes down resemble sentences that occur in my book. Now in my book the sentences are either part of a context that contains qualifications, or they describe views I do not hold. Read with these qualifications or intentions in mind, they correctly express my argument. Gellner does not consider the qualifications and he proceeds as if I were stating my own opinions throughout. The *prima facie* correctness therefore hides some considerable errors.

Take sentence (1) in his 'spinal column'—'the actual history of science shows that the real advances of knowledge contradict all available methodologies' (p. 333)—this is supposed to be a thesis that is held, or insinuated in my book. According to Gellner it is 'clearly the nucleus from which the whole thing grew' (p. 333). To the unwary reader the statement suggests (*a*) that I claim to know the truth of some historical facts and generalisations, (*b*) that I presume to possess insight into the even more difficult problems as to what counts as an advance of knowledge and (*c*) that I refute norms by facts. This is not an abstract possibility. Gellner himself ascribes to me claim (*a*) (p. 337), uses the ascription to accuse me of incoherence (p. 337) and explains my confidence in the face of such incoherence by calling my argument a 'game which [I] cannot lose' (p. 334). But (1) interpreted as entailing (*a*), (*b*), (*c*) is not a thesis I hold. I do not say that methodologies fail because they are merely contradicted by facts—arguments of this kind have been shown to be of questionable value long ago—I say they fail because applied in the circumstances enumerated in my case studies they would have hindered progress. Nor do I claim to possess special knowledge as to what constitutes progress (Feyerabend [1975], p. 27), I simply take the cue from my opponents. *They* prefer Galileo to Aristotle. *They* say that the transition Aristotle → Galileo is a step in the right direction. I only add that this step not only *was not achieved*, but *could not have been achieved* with the methods favoured by them. But does not this argument involve highly complex statements concerning facts, tendencies, physical and historical possibilities? Of course it does, but note that I am not committed to asserting their truth as Gellner surmises. My aim is not to establish the truth of propositions, my aim is to make my opponent change his mind. To achieve this I provide him with statements such as 'no single theory ever agrees with all the known facts in its domain' (Feyerabend [1975], p. 55). I use such statements because I assume that being a rationalist he will be affected by them in a predictable way. He will compare them with what he regards as relevant evidence, for example, he will look up records of experiments. This activity, combined with his rationalistic ideology will in the end cause him 'to accept them as true' (this is how *he* will describe the matter) and so he will perceive a difficulty for some of his favourite methodologies. But don't I now even make more far reaching assumptions about the minds of people, the structure of records, the changes which occur in the former when they are confronted with the latter? Quite right, but these assumptions are not part of my argument with the reader. They are part of an argument I carry on with myself and which concerns the efficiency of my persuasion. The structure of this latter discourse is of no concern for the rationalist who, after all, insists on separating the 'objective content' of an argument from its 'motivation'. All he *needs* to consider, all he is *permitted* to consider is how the statements surrounding the case studies in my book are related to each other and to the historical material and whether they can be read as an argument in *his* sense. I admit that my procedure succeeds by *manipulating* the rationalist but note that I manipulate him in a way in which he *wants* to be manipulated and constantly manipulates *others*: I provide him with material which interpreted in accordance with the rationalistic code creates difficulties for views he holds. Do *I* have to interpret the material as he does? Do *I* have to 'take it seriously'? Certainly not, for the motivation behind an argument does not affect its rationality and is therefore not subjected to any restriction.

Gellner's (2) and (3) and the reasons he says I give for them are equally inadequate. I would not accept the 'this shows' (Gellner [1975], p. 333) for I know that we may 'improve these methodologies' (p. 334); I would not accept the 'all', especially as I have formulated what I believe to be sensible methodological suggestions[1] and argue only against *universal* methods which abstract both from the content of a theory and from the context of the debate (Feyerabend [1975], p. 295, 'fixed and universal rules').[2] Nor would I ever presume to

---

[1] For the use of *ad hoc* hypotheses see Feyerabend [1975], pp. 178 and 97; of a plurality of theories p. 41; of counter induction, chapter 6; 'backward movements', p. 153; connections with influential ideologies p. 193; or other refuted theories p. 142; use of political force to revive theories that are 'scientifically untenable' p. 50; skipping over difficulties, appendix 2—and so on. *Cf.* also my comparison of logical and anthropological procedures for the discovery of methodological rules p. 260.

[2] Gellner's reference to the rules for solving quadratic equations (p. 334) is therefore beside the point (and the example is much more complex than he seems to assume). Furthermore, introducing what he calls his 'own view of the matter' just shows how little he has understood mine: there is no difference between the two.

legislate for scientists or, for that matter, for anyone else as Gellner implies in (5) and (6). I did this in earlier papers when I was younger, more ignorant, more pushy and considerably more conceited.[1] At that time my arguments for proliferation were indeed designed to show that a monistic life is not worth living and they urged everyone to think, feel, live through a competition of alternatives. Today the same arguments are offered with a very different *purpose* in mind, and they lead to a very different *result*.[2] Scientists and rationalists have by now almost succeeded in making their views the basis of Western Democracy. They concede, though with extremely bad grace, that other ideas may be *heard*, but they would not permit them a role in the planning and the completion of fundamental institutions such as law, education, economics. Democratic principles as they are practiced today are therefore incompatible with the undisturbed existence, development, growth of special cultures. A rational-liberal democracy cannot contain a Hopi culture in the full sense of the word. It cannot contain a Black culture in the full sense of the word. It cannot contain a Jewish culture in the full sense of the word. It can contain these cultures only as *secondary grafts* on a basic structure that is constituted by an unholy alliance of science, rationalism (and capitalism). All attempts to revive traditions that were pushed aside and eliminated in the course of the expansion of Western culture and to make them the basis of existence for special groups run into an impenetrable stone wall of rationalistic phrases and prejudices. I try to show that there are no arguments to support this wall and that some principles implicit in science definitely favour its removal.[3] There is no attempt on my

---

My argument is directed against philosophical methodologies that claim to be 'rational' as well as universally valid and that do not arise from an analysis of concrete circumstances. It says that the rules of such methodologies will hinder progress and that the only rule that is both general and does not interfere with scientific progress is as empty and useless as 'anything goes'. It also says that the methodological decisions we make in concrete circumstances *constitute* rationality rather than being guided by pre-existing canons of rationality and cannot in any way be restricted in advance, not even by the rules of formal logic.

[1] Gellner does Popper an injustice when connecting this immature earlier work with the 'Popperite movement' (p. 332). True, there are some acknowledgments to Popper in these papers, but they are friendly gestures, not historical statements (I also mention my girl friends). True, some points sound quite Popperian to somebody who has only read Popper, but they stem from Mill, Mach, Boltzmann, Duhem and above all, Wittgenstein. True, in my book I occasionally make fun of Popper (I 'lash out' says Gellner—p. 332—who does not seem to be able to distinguish mockery from aggression), but only to tease Lakatos (Feyerabend [1975], p. 8) who was supposed to reply to my book and who was overly impressed by Popper, and not because of 'excessive demands for conformity and involvement made by the master' (*ibid.*, 'master' indeed!).

[2] The change is due to a conversation that occurred at Professor von Weizsaecker's seminar in Hamburg 1965: Professor von Weizsaecker gave a detailed account of the Copenhagen interpretation and showed how it could be applied to specific problems. I pressed the point that hidden variable theories are needed to increase the empirical content of the orthodox view when I suddenly realised how barren such an attitude is in the face of concrete research.

[3] Gellner says that the social consequences of rationalism (or irrationalism) are 'tangential' to my main concerns (p. 339). *The very opposite is the case.* For me democracy, the right of people to arrange their lives as they see fit comes first, 'rationality', 'truth' and all those other inventions of our intellectuals come second. This, incidentally, is the main reason why I prefer Mill to Popper and why I have nothing but contempt for the sham-modesty of our critical rationalists who fall over themselves in their concern for 'freedom' or

part to show 'that an extreme form of relativism is *valid*' (p. 336), I do not try to *justify* 'the autonomy of every mood, every caprice, and every individual' (*ibid.*), I merely argue that *the path to relativism has not yet been closed by reason* so that the rationalist cannot object to anyone entering it. Of course, I have considerable *sympathy* for this path and I think it is the path of growth and freedom, but that is a different story.

More specifically, the situation is as follows. I do not show that proliferation *should be used*, I only show that the rationalist *cannot exclude it*. And this point is not made negatively, by showing how existing objections break down but *via* an argument that derives proliferation from the monist's own ideology. The argument is in two parts, one based on science, the other on the relation between scientific and non scientific ideologies. The argument from science says that proliferation follows from the scientist's own demand for high empirical content (Feyerabend [1975], pp. 41 and 47). I do not accept the demand which is only one of the many ways to bring order into our beliefs (*ibid.* p. 204) and so I do not argue for its implications. What I do say is that scientists who are in favour of high empirical content are committed to proliferation as well and therefore cannot reject it.[1] The argument from the existence of incommensurable ideologies says (*a*) that their comparison does not involve content and so cannot be in terms of truth *vs.* falsehood, except rhetorically (ch. 17),[2] (*b*) that every ideology

---

'Open Society' but start erecting obstacles whenever people want to live in accordance with the traditions of their forefathers. *Cf.* Agassi's review of my book and my reply in *Philosophia* (Jerusalem), January 1976 as well as my introduction to the third edition of *Against Method.*

[1] I explicitly state that 'my intention is not to replace one set of general rules by another such set: my intention is, rather, to convince the reader that *all methodologies, even the most obvious ones, have their limits*' (Feyerabend [1975], p. 32 my italics)—which does not stop Gellner from presenting 'pointless proliferation of viewpoints' (p. 340) as a *positive* doctrine of mine and explaining recalcitrant passages either by some hidden incoherence on my part (p. 333) or by my 'clowning' (p. 338).

At first sight this looks as if Gellner had made an interesting contribution to the art of rhetorical argument: Assume you are supposed to review a book the major part of which is beyond your comprehension (p. 334). You concentrate on the 'rest of the book' (*ibid.*), say that 'it is of some interest' (*ibid.*) and extract a series of theses from it. You also present the theses in a systematic way, you arrange them in a 'spinal column' (p. 333) and add arguments to show that you are proceeding in a fair and rational manner. If the arguments or the theses contradict statements found in the book, you accuse the author of incoherence. If the author is not your run-of-the-mill academic square but makes a joke now and then you can explain the conflict by his 'clowning' as well. So you can have your cake—you don't have to understand everything you read—and eat it, too—you can write a vigorous, decisive, witty review. Unfortunately Gellner is only dimly aware of what he is doing. Most of the time he thinks he is producing a straightforward rational criticism (*passim*). So we cannot praise him for his rhetorical acumen but are reduced to noting his inability to understand what he reads.

[2] It is therefore not correct to describe my view as implying that 'almost everything might contain some truth' (p. 335). Gellner's 'Epistemological theories . . . give us some insight in how to choose between whole styles of thought' (p. 336) does not solve the conundrum either for every self respecting 'style of thought' will of course have its own epistemology (*cf.* Feyerabend [1975], p. 246). The principle, however, that a 'culture which subjects its cognitive capital to testing by arbiters *who are not under its own control*' is superior to one 'which does not do so' (p. 336) would prefer cultures with oracles to cultures with scientific experiments for the latter are on the whole much more strictly controlled than the former.

possesses methods of its own and that the comparative evaluation of methods has not even started. All we have is the dogmatic belief in the excellence of the 'methods of science' (where everyone has different ideas as to what these methods are). But, (*c*), non-scientific views and methods, far from being complete failures have led to amazing discoveries in the past, they are often better than the corresponding scientific views and have better results (*cf.* Feyerabend [1975] pp. 49 ff.).[1] Taking all these arguments together I conclude that a person who wants to introduce unusual views, methods, forms of life or who wants to revive such reviews, methods, forms of life *does not need to hesitate* for reason has not yet succeeded in putting any obstacles in his path and scientific reason even urges him to increase the number of alternatives. The only obstacles he is going to meet are prejudice and conceit.

Let us stay a little longer with this matter of proliferation to get a good look at Gellner the reviewer in action. We have seen that Gellner misjudges the *role* of proliferation in my arguments. He does not understand its *consequences* either. He chides me for the 'welcome admission' (p. 339) that technology cannot be had without scientists. To start with, there is no such admission on my part. I address people who fear that the separation of state and science will lead to a breakdown of health, public transportation, radio, T.V. and so on because—and this is *their* reason, not mine—technology cannot exist without scientists (Feyerabend [1975], p. 299). Trying to alleviate this fear I could either deny the reason, *i.e.* I could argue that technology does not need closed societies of highly qualified experts to succeed—this I do on page 307, though rather briefly—or I could give an answer that leaves the reason untouched—and this I do on page 299. Assuming that my reader can follow an argument without being constantly reminded of its presuppositions I state the view of my opponent and my own views side by side, as in a dialogue, but without making the sides explicit. Page 299, *e.g.*, means: *Opponent*: but will not a separation of state and science lead to a breakdown of technology? *I*: you seem to belive that technology without experts is impossible, I am rather doubtful about this assertion, but let us take it for granted. Then you must realise that there will always be people who prefer being scientists . . . and so on. Gellner conflates the opponent's thesis and my own reply, turns the conglomerate into a single view, ascribes the view to me, analyses it, and triumphantly exposes its incoherence. And as he starts running different contexts into each other whenever the argument becomes a little complicated, he has now another, and most efficient method of finding inconsistencies in my book. But the 'plot' (p. 338) he unveils in this manner is but a reflection of his own simpleminded reading habits: he understands 'the cat is on the mat'; he may still understand, though with some effort 'Joe says that the cat is on the mat' but 'the cat is on the mat—do you really believe this? I don't' shows to him that the author says that the cat both is and is not on the mat and thus defends incoherence. This is Gellner's third 'contribution' to the art of rhetorical argument.[1]

---

[1] Gellner is 'sceptic[al] about [such] truly amazing achievement[s]' (Gellner [1975], p. 242) and this is understandable, for he does not know the literature. What he does know is that most of his readers share his scepticism and will be impressed by a statement of it. Had he also known that they are sceptical because they are illiterate then we could have congratulated him on his elegant use of a fine rhetorical principle: if your opponent assumes matters your readers are not likely to have heard about then score a point by acting as if these matters did not exist and as if your opponent was talking through his hat. But Gellner thinks his information is complete which means: we are not dealing with rhetorical sophistication but with ignorance, pure and simple.

To sum up: though I am personally in favour of a plurality of ideas, methods, forms of life, I have not tried to *support* this belief by argument. My arguments are rather of a negative kind, they show that reason and science *cannot exclude* such a plurality. Neither reason nor science are strong enough to impose restrictions on democracy and to prevent people from introducing their own most cherished traditions into it. (Another result is that rationalists have not yet succeeded in defeating scepticism—all views are *equally good*—or its natural extension—*any* evaluation of theories and forms of life is acceptable (*cf.* Feyerabend [1975], p. 189).[2]

This is perhaps the place to make some remarks on the meaning of the phrase 'Anything goes'. Many readers, Gellner included, understand the phrase as recommending chaos and arbitrariness. This is not the case. I do not object to rules, standards, arguments, I only object to rules, standards, arguments *of a certain kind.* I object to rules, standards, arguments which are *general*, and *independent of the situation in which they are applied.* To my mind the idea of a methodology that contains such rules is as laughable as the idea of a measuring instrument that measures any magnitude in all circumstances. We know that different magnitudes have to be measured in different ways, that the methods of measurement have to change with temperature, the material, the stress that occurs during the measurement and we also know that there will always be circumstances when we must invent entirely new instruments. In the same way we have to revise our *intellectual* instruments as we go along. We may use old familiar procedures in old and familiar circumstances, we may have to invent entirely new procedures, new standards, a new logic in new domains. The criteria which guide the invention of new physical instruments will of course depend on the new laws we find in the new domains, and so they cannot be judged by the customary standards of resilience and sensitivity. In the same way the criteria which guide the invention of new standards, new rules of procedure, of a new logic will depend on the nature of the new domain, and so again they cannot be judged by the standards, or the logic already in existence (*cf.* what I say about the cosmological criticism of methodological rules on p. 206). From the point of view of the accepted standards the research that leads to the new standards is of course arbitrary, irrational, but it is not at all arbitrary and irrational when judged by the new standards, relative to the new domain. There is no rule that is valid under all circumstances just as there is no measuring instrument that measures everything and in all circumstances, but it is possible to construct such a rule in a purely formal manner: it is the rule 'anything goes'. Does this mean that there will not be a theory of knowledge? Not at all. There will be lots of rules of thumb with practical advice about their limits and their judicious application inside the limits, but there will not be any general principles. The theory of knowledge I envisage has therefore much in common with the science of rhetoric; it has very little in common with the science of logic, in its present shape. All this means, of course, is that methodology is taken out of the hands of armchair philosophers and logicians and given back to the researchers themselves. It is they who advance into the unknown, it is they who must decide which (physical or mental) instruments to choose.

---

[1] Typical example: on p. 21, note 12 I defend the pacifism of the Dadaists and say that I am against violence. On p. 187 I say that *political*, or *eschatological* anarchism regards violence as necessary. Gellner (p. 340) runs the two passages into one and says that I 'incoherently conflate' (beautiful words!) the 'mystique of violence' with a 'pacifist no-fly-hurting posture' which in turn is combined with 'cognitive/productive parasitism'. Well roared pussycat—but don't you think you should have read the text a little more carefully or let somebody else explain it to you in case you can't read? The text says that violence is necessary *according to political anarchism* and adds that political anarchism is a doctrine I reject. The very first sentence of the book calls political anarchism 'not the most attractive political philosophy' (p. 17) and on p. 189 I again distinguish my views from political anarchism, just to be on the safe side. All in vain.

There is a second point to be made about the 'admission' Gellner ascribes to me: is it not contrary to the idea of proliferation. Proliferation does not mean that scientists are excluded, or that statements such as 'we need scientists' or 'Lysenko failed' (p. 341) are *removed* from the domain of debate, it means that statements denying them, or making fun of them are *admitted* in the hope that this will be of advantage. With liberalism the situation is exactly the same. Gellner chides me for explaining the difference between Popper and Mill by reference to Popper's puritanism. 'My own liberalism' he says with pride (p. 332) 'is such that I hold that even Puritans are not debarred from truth'. I did not say they were, I said that Popper's liberalism is different from Mill's and that Puritanism is the explanation. Nor does one become illiberal when denying truth to a Puritan. Liberalism is an *institutional* doctrine, not a doctrine about *individual beliefs.* It does not regulate individual beliefs, it says that nothing may be excluded from debate. Thus, being a liberal, I do not have to love Puritans, I do not even have to admit that they have a chance of finding the truth, all I am required to do is to let them have their say and not to stop them from doing so by *institutional* means (I may try to persuade them privately). It is a strange liberalism indeed that assumes that everyone has a chance of finding truth (and would therefore frown on a defender of falsehood).

[2] Gellner ascribes the extension to 'temperamental exuberance' (p. 335) while it is in fact the result of an application of the sceptic's method (of balancing each judgment by its opposite) to his own basic principle (all views are equally good). Here as elsewhere Gellner is quick to explain positions by psychology when they are in fact the results of argument.

(*ii*) Rationalists cannot rationally exclude myth and ancient traditions from the basic fabric of democracy. Yet they push them aside, using sophistry, pressure tactics, dogmatic pronouncements many of which they regard as arguments, and *present* in the form of arguments. Such pseudoreasoning can either be exposed by scholarly analysis, or it can be held up to ridicule. I chose the latter path, partly because I had given arguments where they were necessary, partly because I could not see myself solemnly taking apart these products of conceit and pomposity. Gellner neither likes my procedure, nor does he understand its function. He thinks I use it as a 'criticism evading ploy' (p. 338) when in fact I apply it in a domain where the opponent does much huffing and puffing but is no longer engaged in a rational debate. Having decided to disregard 'the extensive parts [of my book] which argue [this] viewpoint' (p. 333) Gellner was without guidance when entering the domain; being taken in as much as everyone else he did not discover its limits and so, quite naturally, he is upset at what he thinks is unfair and irrational treatment of people 'who ask questions about knowledge in good faith' (p. 342). But the trouble is that this 'good faith' is a faith in principles that are far beyond the reach of argument and are accepted only because of the rationalist's say-so; the trouble is that it is a faith in principles that belong to the theology of rationalism.

Gellner also objects to the use of ridicule and frivolity. '[E]ncouraged by the spirit of the times' he writes (p. 334) both have been 'allowed to intrude within the covers of the book.' They have been 'allowed to intrude'—this means that ridicule and frivolity ***are*** there, but ***should not*** be there. Why not? Presumably because they should not occur in books of a certain kind, for example they should not occur in scholarly books. Now, first of all, what little bird has told Professor Gellner that I intended to write a scholarly treatise? In my dedication I make it clear that my book was conceived as a *letter* (p. 7) to Lakatos and that its style would be that of a letter. (Also, I am not a scholar, and have no wish to be a scholar.[1]) Secondly, why should scholarly books be dry, impersonal, lacking in frivolity? In the nineteenth century and earlier, scholarly debates were very lively, the number of insults about the same as the number of footnotes. Dictionaries of recondite languages (medieval Latin/English; Sanscrit/English) used racy equivalents, introductions to important editions teemed with ambiguous insinuations. Then gradually, a more measured tone set in, people became more solemn, they frowned on levity and behaved as if they were playing parts in a strange and highly formalised play. Language became as colourless and as indistinct as the business suit which is now worn by everyone, by the scholar, by the businessman, by the professional killer. Being accustomed to a dry and impersonal style the reader is disturbed by every deviation from the dreary norm and sees in it an obvious sign of arrogance and aggression; viewing authority with almost religious awe he gets into a frenzy when he sees someone pluck the beard of his favourite prophet. This, my dear Professor Gellner, is the 'spirit of the times' and not the attempt by a few outsiders to restore older and less formalised ways of writing. I do not know the reasons for the change, for there was never an argument, I do not see any advantage in it, and so I do not see why I should accept it as a *fait accompli*. Let us now have a look at the explanation Gellner offers for my dissent.

(*iii*) According to Gellner I 'conflate incoherently' a 'mystique of violence' with a 'pacifist no-fly-hurting posture' to which I add 'cognitive/productive parasitism' (p. 340).

We have seen how the first part of this charge arises[2]: Gellner 'conflates' passages which express my own views with passages describing the views of others. The incoherence is in his reading, it is not in my text. The second part of the charge is a puzzle to me and I can only explain it by some scientistic tendencies on the part of Gellner, or by some amazing lack of reading skills. On page 300, which Gellner quotes, though incompletely, I say that scientists may have some interesting ideas and gadgets to offer, that we should listen to their ideas and use their gadgets but without permitting them to build society in their own image, *e.g.* without permitting them to become the masters of education: there should be a separation of state and science just as there is now a separation between state and church. The reason for the separation is simple: every profession has an ideology and a drive for power that goes far beyond its achievements and it is the task of democracy to keep this ideology and this drive under control. Science is here no different from other institutions as can be seen from the attitude of official medicine towards unusual ideas that have not passed through its own channels (and note that the comparative efficiency of these channels has never been examined in the past and that research that is being carried out today has revealed appalling shortcomings). Calling such a procedure 'cognitive parasitism' is as sensible as calling parasites all those astronomers who consult ancient records but without adopting the theology that played an essential part in their construction and interpretation. And as regards the 'productive' side of this 'parasitism' I need only repeat that scientists will of course be amply rewarded for their services[1]—which is more than is granted to the taxpayer who is supposed to finance scientific research but without any assurance that his needs will be taken into account.[2]

---

[1] Gellner says history and philosophy of science are 'area[s] of [my] professional expertise'. They are not as every historian and philosopher of science will be most happy to confirm.

[2] See p. 387, note 1.

Having stated his version of my views, Gellner starts classifying them. How does he proceed? He has heard that I am now in Berkeley and that some people in Berkeley some time ago preached peace but were attracted by violence. 'Putting two and two together' he calls my 'doctrines about violence' (which, as we have seen, are doctrines I describe and reject) 'Californian' which will hardly please Ronald Reagan and his numerous followers all over the state, from Los Angeles, Orange County all the way up to Goose Lake and does injustice to the many fine and upstanding revolutionaries that arose at the LSE. Next Gellner remembers that I was born in Vienna and he has also the idea, no doubt from some American movies, that the Viennese like to live a relaxed life. Again 'putting two and two together' he calls some of my suggestions 'characteristically Viennese' (p. 332) which is just as sensible as calling Popper's dogmatism 'Popish conceit' on the grounds that Popper is from Vienna and that Vienna is full of Catholics. One sometimes wonders whether Gellner took these explanations really seriously or whether they are not rather attempts to use rhetoric when reason fails him. A rhetorician, of course, would know that invective is successful only if its main elements are based on fact and if it does not offend the intelligence of the reader. Neither condition is satisfied in Gellner's case. So we are again[1] reduced to noting that what at first sight seemed such a promising piece of rhetorical exuberence is nothing but a simple failure of comprehension and rationality.

---

[1] Gellner omits this passage from his quotation and replaces it by dots so maybe, he is not such a bad reader after all.

[2] The charge of cognitive/productive parasitism stands the actual situation on its head. What is a parasite? A parasite is a man or a woman who gets something for nothing. Today many scientists and intellectuals are parasites in precisely that sense. They get something—large salaries, expensive playpens—for nothing. For let us not forget that only a fraction of the research and the teaching that is being carried out at state universities and other tax-supported institutions such as the National Science Foundation benefits the community at large, or is even conceived with the idea of such benefits. Even research that seems eminently practical is conducted in a way that decreases the chance of speedy practical results: one does not explore successful procedures that are theoretically opaque, one prefers an approach that gives 'understanding' where the criteria of understanding are defined by the researchers themselves, *e.g.* cancer research. Alternative procedures are rejected without examination not because they are faulty, but because they clash with the equally unexamined beliefs of one's own sect. The attitude has deplorable consequences in education. Valuable traditions are eliminated, the lives of people are impoverished not because the traditions have been shown to be inadequate but because they don't agree with the basic assumptions of science and because scientists have now the power to impose their ideology on almost everyone. Thus scientists and intellectuals are not just parasites of the pocketbook, they are parasites of the mind as well and they will continue in their path unless democracy puts them in their place. This is what I suggest. I suggest that there be a careful examination of the way in which scientists use public money and of the doctrines they impose upon the young. I suggest that promising research be properly rewarded—but the underlying ideology will not automatically become part of basic education (the ideology of prison guards may be excellent for keeping prisoners in their place, but it may be entirely unsuited as a basis for general education). Needless to say I welcome the Baumann amendment which recommends congressional veto power over the 14,000 odd grants the National Science Foundation awards every year. It is a *very* small step, but a step in the right direction. Scientists were upset when the amendment was passed by the house of representatives and the director of the National Academy spoke darkly of totalitarian tendencies. The well paid learned gentleman did not seem to realise that totalitarianism means direction of the many by the few while the Baumann amendment goes in the opposite direction: it suggests examining what the few are doing with the millions of public money that are put at their disposal in the hope that the public will eventually profit from such generosity. Considering the narcissistic chauvinism of science such an examination would seem to be more than reasonable. It takes a sociologist to describe the procedure as 'cognitive/productive parasitism', but it is understandable why Gellner reacts in this way: realisation of my plan means the end of a comfortable life for him.

This brings me to the last point of my reply, *viz*. Gellner's attempt to defend Lakatos. Lakatos, Gellner says, 'observed the highest standards of rigour, lucidity and responsibility' both in 'his writings and [in his] lectures' (p. 332). Poor Imre! If one thing is certain it is that the standards of Lakatos were very different from those of Gellner. Lakatos was not at all averse to tricks, frivolity, ridicule in his lectures, he never descended to the *tierischer Ernst* which is Gellner's basic attitude, despite all his attempts at levity, and even in his writings he more than once left the path of rational argument to take a well aimed swipe at his opponent. On the other hand, Lakatos could read, and he had only contempt for the types of explanation I have just discussed and which Gellner uses not just in his review, but in other works as well, for example in *Words and Things*. Nor was it necessary to warn the reader of my dedication (p. 331). Imre Lakatos, whom I asked permission accepted it with glee; he knew that it was a joking reference to chapter 16 that deals with his views and he looked forward to writing his reply. This reply, I am sure, would not have consisted in the bald statement that 'the claim that Lakatos' position was an 'anarchism in disguise' is totally unfounded (p. 331), for he paid attention to my arguments and thought that I had made a good case for this claim. Read it as you will, Gellner's praise of Lakatos and his attempt to defend him from me is an undeserved insult to the memory of a great scholar and a marvellous human being.

PAUL FEYERABEND
University of California, Berkeley

REFERENCES

FEYERABEND, P. [1975]: *Against Method*, New Left Books.

GELLNER, E. [1975]: 'Beyond Truth and Falsity', *British Journal for the Philosophy of Science*, **26**, pp. 331–342.

---

[1] *Cf.* n. 1 p. 385 and n. 1 p. 386 and text to n. 1 p. 387.

ON THE PROBABILITY OF LAWS BEING ZERO

The problem of whether it is meaningful to assign a probability to the validity of a universal contingent law with an infinite domain and, if so, whether this probability can be anything but zero has a large literature. After briefly looking at some of the contributions to this problem, I present an axiomatisation of probability statements which leads directly to the conclusion that such laws must have zero probability. Although some attempt is made to argue for the plausibility of the axioms, it may be that a better approach to the results below is to assume that laws exist which do have nonzero probability and hence that at least one of the axioms is invalid.

Popper[1] presents a number of arguments for the case that if *a* is any law (to be taken herein, unless otherwise specified, as a universal contingent law with an infinite domain) and *b* is any empirical evidence for *a*, then the logical probability of *a* (being always valid) is necessarily zero. On the other hand, Howson has recently given a criticism of Popper's arguments and then presented a case for the existence of laws with a nonzero probability. However, he concludes with the remark that the 'reasonable attitude to law sentences in infinite domains is to be extremely skeptical of their truth'.[2]

Turning to a recent book by Hesse ([1974]), the author writes that:

> it is reasonable to have zero degree of belief in any universal generalisation in an infinite domain, as this is actually expressed at any given stage of science.[3]

However, on a first examination this does not seem to be consistent with her earlier statement that:

> the use of ideal geometric models in physical theory makes some precise mathematical laws significant, and incidentally gives them nonzero probability, because . . . only a finite number of ideal geometric shapes may figure in the qualitative language.[4]

(The particular example used is the inverse square law of gravitation.) It would seem that the crucial point for Hesse is to be able to distinguish between 'universal generalizations' and 'precise mathematical laws'. Perhaps it is proper to compare this latter case with those types of laws which, as Eddington puts it, 'correspond to *a priori* knowledge'[5] and as such must be certain since they are 'a consequence of the conceptual frame of thought into which our observational knowledge is forced by our method of formulating it'.[6] A difference is, however, that Eddington puts 'all the laws of nature that are usually classed as fundamental'[7] into this category.

Thus we have examples of authors separately arguing that the probability of laws (*i*) must always be zero, (*ii*) may be nonzero, and (*iii*) must always be one. (Strictly speaking, the examples supporting the third possibility are not admissible

[1] Popper [1959], Appendix *vii.

[2] Howson [1973], p. 163.

[3] Hesse [1974], p. 193.

[4] *Ibid.*, p. 185.

[5] Eddington [1939], p. 57.

[6] *Ibid.*, p. 104.

[7] *Ibid.*, p. 57.

since the central thrust of Eddington's philosophy is that the laws referred to are not contingent, but are necessary consequences of the enveloping theory. Nevertheless, it is interesting to note that the same law statement could be viewed by Popper to have zero probability, by Hesse (and perhaps Howson) to have a nonzero probability and by Eddington to be certain.)

This completes the *mise en scène* and so we now turn to the promised axiomatisation. Only simple enumeration will be considered and so it will be supposed that empirical data from some source or sources is continually being coded into 0's and 1's. (A justification for this is that if the incoming data cannot be formalised in this manner, then it would be highly incongruous to talk of a numerical probability of the validity of a law predicting this data.) Two of the types of coding that are possible are illustrated in the following examples.

(*i*) The distance between the earth and moon is measured to the nearest one hundred kilometres every minute, divided by one hundred, and then expressed in binary notation.

(*ii*) All data regarding experiments on Einsteinian mechanics are sent to a central office where they are assigned 0 if the data agrees with Einsteinian predictions within 50 per cent or 1 otherwise. (In this case some subjectivity can be allowed to enter by permitting the operator to assign any finite string of 0's to a supporting experiment, the length of the string being a subjective estimate of the significance of the experiment.)

Such a sequence will be called a *data sequence*.

A *deterministic law* is an effective rule for generating in succession a sequence of 0's and 1's in such a way that at any stage the next outcome can always be predicted; the sequence generated will be called a *law sequence*. Given a data sequence D, a law sequence L is said to be *valid* (with respect to D) if D and L are the same. Also D is said to *support* L to a degree n if the first n terms of D and L agree.

Let N denote the set of natural numbers and for any k in N define P_k as the set of finite sequences

$$(D_1,L_1,n_1;\ D_2,L_2,n_2;\ \ldots\ ;\ D_k,L_k,n_k)$$

where: $n_1, \ldots, n_k \in N$; $D_1, \ldots, D_k$ are data sequences; $L_1, \ldots, L_k$ are laws; and each D_j supports L_j to a degree n_j. Now define P as the union of all the P_k. If D_1 and D_2 are data sequences, define $D_{1,2}$ as the data sequence obtained by taking in succession the first element of D_1, the first element of D_2, the second element of D_1, *etc.* Similarly, define $L_{1,2}$ as the law sequence derived from L_1 and L_2 by taking alternately from L_1 and L_2, beginning with the first element of L_1. Finally suppose that we have a real-valued map p on P with the following properties:

*A*1 (Axiom of probability) (*a*) For any $x \in P$, $0 \leqslant p(x) < 1$. (*b*) If D supports L to a degree $n+1$, then

$$p(D,L,n) \leqslant p(D,L,n+1).$$

*A*2 (Axiom of weak independence) If $(D_1,L_1,n) \in P_1$ and $p(D_1,L_1,n) > 0$, then there exists $(D_2,L_2,n) \in P_1$ with

$$p(D_1,L_1,n;\ D_2,L_2,n) < p(D_1,L_1,n).$$

Furthermore, in this case.

$$p(D_{1,2},L_{1,2},2n) = p(D_1,L_1,n;\ D_2,L_2,n).$$

*A*3 (Axiom of objectivity) If (D_1,L_1,n) and (D_2,L_2,n) belong to P_1, then

$$P(D_1,L_1,n) = p(D_2,L_2,n).$$

If $(D_1,L_1,n_1;\ \ldots;\ D_k,L_k,n_k)$ belongs to P, then its image under P should be thought of as the probability that $L_1, \ldots, L_k$ are valid laws (predicting $D_1, \ldots, D_k$ respectively) given that D_j supports L_j to a degree n_j for $j = 1,2, \ldots, k$. Then *A*1 (*a*) merely states that no matter how much supporting data we have, the validity of any law is never certain, while *A*1 (*b*) states that the more supporting data we have, the more likely is the validity of the law . . . at least it is not less likely.

Turning to *A*2, if $(D_1,L_1,n) \in P$ and if there exists $(D_2,L_2,n) \in P$ such that the validities of L_1 and L_2 are independent, then

$$p(D_1,L_1,n;\ D_2,L_2,n) = p(D_1,L_1,n).p(D_2,L_2,n).$$

Thus, in view of *A*1 (*a*), if $p(D_1,L_1,n) > 0$,

$$p(D_1,L_1,n;\ D_2,L_2,n) < p(D_1,L_1,n).$$

The first part of axiom *A*2 proceeds straight to this inequality; the interpretation of the remainder of the axiom is obvious. The final axiom *A*3 can be interpreted as saying that the probability of any law being valid is dependent *only* upon the number of units of supporting empirical data. For example, if it is felt that for some reason (such as intuition or more satisfying theoretics) one law is more likely to be valid than another when they are both supposed to be predicting the same empirical data, then the opportunity is open to obtain more data until one or other of the laws is disqualified. This would, of course, happen not later than at the first place where the law sequences differed. But until this place is reached, both laws would have to be considered as being equally likely to be valid.

At first glance this axiom may seem too stringent or counter-intuitive to be acceptable. However, these objections are weakened if we only consider data sequences formed in the manner of example (*ii*) above.[1] For in this case the law sequence always consists entirely of 0's and the opportunity is open to assign more than one 0 in the data sequence to, for example, experiments supporting a more comprehensive theory or simply more substantial experiments. In this way the number of units of confirmation can be increased at a faster rate which would thus push up the probability of the validity of the law sequence as in *A*1 (*b*).

THEOREM. *Whenever L is a law sequence which is supported by a data sequence D to an even degree n, then* $p(D,L,n) = 0$.

---

[1] Notice that any data sequence-law sequence pair can be converted to this form; the new data sequence is formed by putting one or more 0's each time there is agreement between the original data and law sequences while the new law sequence consists entirely of 0's.

Proof. Suppose that $(D_1,L_1,n) \in P$ where n is even. It will be shown that the assumption '$p(D_1,L_1,n) \neq 0$' leads to a contradiction. By *A*2 there exists $(D_2,L_2,n) \in P$ such that

$$P(D_{1,2},L_{1,2},n) < p(D_1,L_1,n/2). \qquad (1)$$

(Recall that if $(D_1,L_1,n) \in P$, then *a fortiori* $(D_1,L_1,n/2) \in P$.) By *A*3 we can replace $D_{1,2}$ by D_1 and $L_{1,2}$ by L_1 in (1) to obtain

$$p(D_1,L_1,n) < p(D_1,L_1,n/2). \qquad (2)$$

But this contradicts *A*1 (*b*) and so the only possibility remaining is that $p(D_1,L_1,n)$ is zero. This possibility is clearly compatible with the axioms.

Surprisingly, it seems that a proof of '$p(D,L,n) = 0$' for n odd cannot be obtained from the axioms as they stand although it can be proved by adding an extra axiom. For now I will be content with the observation that a trivial consequence of the above theorem is that if $(D,L,n) \in P$, then $p(D,L,k) = 0$ for all natural numbers k up to and including the largest even integer not exceeding n.

A final comment is that it is not difficult to find other axiomatisations which lead to the result of the above theorem. For example, given (D,L,n) in P, assume (*i*) that there are an infinite number of law sequences L' supported by D to a degree n and (*ii*) that the sum of the $p(D,L',n)$ over these sequences does not exceed unity. An application of axiom A3 leads immediately to the conclusion that $p(D,L,n) = 0$. On the other hand, if the sum in (*ii*) is assumed to be unity, then there must exist a law sequence L_1 supported by D to a degree n which has $p(D,L_1,n)$ nonzero.

JOHN F. PRICE
University of New South Wales

REFERENCES

Popper, K. R. [1934]: *Logik der Forschung.*

Popper, K. R. [1959]: *The Logic of Scientific Discovery* (expanded English translation of Popper [1934]).

Howson, C. [1973]: 'Must the logical probability of laws be zero?', *British Journal for the Philosophy of Science,* **24**, pp. 153–82.

Hesse, M. B. [1974]: *The Structure of Scientific Inference.*

Eddington, A. E. [1939]: *The Philosophy of Physical Science.*

FRIEDMAN'S CRITERION FOR SIMPLICITY[*]

In a recent series of articles,[1] Kenneth Friedman articulates a formal criterion of comparative empirical simplicity of predicates, claiming that it is superior to earlier criteria by being independent of the logical strength and syntactical form of the predicates involved. Using this criterion, Friedman proposes answers to Goodman's 'new riddle of induction' and Hume's problem of the justification of induction. In this paper, I point out some defects in Friedman's argument to show that, by his criterion, 'green' is simpler than 'grue'. I go on to show that, given his definitions, ***no predicate is simpler than any other!***

---

[*] I wish to thank Charles Chihara, who suggested this topic, for stimulating and encouraging discussions. Also, a letter from Kenneth Friedman was very helpful.

[1] Friedman [1972], [1973], [1975*a*].

According to Friedman, C is ***simpler than*** D iff the set of tests for D (excluding those depending on certain 'parasitic information' $I(CD)$) is a proper subset of the set of tests for C (excluding those depending on $I(DC)$). A *test for a predicate* is a procedure which implements a sufficient condition for the applicability of the predicate (positive test) or a sufficient condition for the inapplicability of the predicate (negative test) or both. For any predicates I and C, a test for C is ***independent of the information*** I iff there is no object x such that the outcome of the test, applied to x, is incompatible (given, I assume, known general facts about the world) with either the applicability or the inapplicability of I to x; otherwise, the test ***depends on*** I. The crucial definition concerns ***information for*** D ***parasitic on*** C $I(CD)$: A predicate I represents such information iff it is of minimal logical strength among the predicates satisfying (*1*) and (*2*) below or a variant thereof obtained by consistent substitution of negations of one or more of I,C, and D in (*1*) or (*2*) or both.[1]

(*1*) $$(x)\{Cx \rightarrow (Ix \leftrightarrow Dx)\}$$

(*2*) $$(K)[(x)\{Kx \rightarrow (Ix \leftrightarrow Dx)\} \rightarrow (x)(Kx \leftrightarrow Cx)]$$

Friedman's argument that 'green' is simpler than 'grue' (in our world) can be summarised as follows[2]:

(*i*) Looking at an object in a good light and seeing whether or not it is green is a test for 'green'. This test depends only on whether or not the object is green (and thus does not depend on any information for 'green' parasitic on 'grue'). Furthermore, this is not a test for 'grue', since the determination as to whether an object is green or not does not implement a sufficient condition for its being grue or for its being nongrue. Thus, there is a nonexcluded test for 'green' which is not a test for 'grue'.

(*ii*) Using 'Ix' as an abbreviation for 'x was first inspected before T.', we can infer that Ix is information for 'grue' parasitic on 'green', for if x is green, then x is grue if and only if Ix, and there is no other predicate (but 'green') with this property. Thus I satisfies (*1*) and (*2*).

(*iii*) A test for 'grue' based on showing an object blue or green must depend on I and is thus excluded as a test for 'grue'.

(*iv*) A test for 'grue' based on showing an object to be nongreen and nonblue may be independent of I (and so presumably not excluded as a test for 'grue'), since a nongreen nonblue object is nongreen and nongrue no matter when first examined. But such a test is also a test for 'green' (and, Friedman would presumably hope, not excluded as a test for 'green').

(*v*) Thus the set of nonexcluded tests for 'grue' is a proper subset of the set of nonexcluded tests for 'green'.

---

[1] This formulation is somewhat different from those appearing in Friedman [1972], [1973]; the modifications are derived from Friedman [1975*b*]. The changes do not require modifications of Friedman's earlier arguments; rather the present version is to represent a more accurate formalisation of the principles used in those arguments. A discussion of Friedman's theory leading to conclusions similar to those of the present article, but based on different versions of axiom (3) and the negation substitution conventions, is given in Manders [1974].

[2] Friedman [1973], p. 371.

This argument contains two defects. Firstly, it is asserted without proof in (*i*) and (*iv*) that certain tests for 'green' are independent of whatever information there may be for 'green' parasitic on 'grue'. This is intuitively plausible, but a careful consideration of the formal definitions carried out below shows that the claim is in fact false. Secondly, claim (*ii*) is false: *Gred* objects are grue if and only if they are first observed before T; so 'green' is not the unique predicate with this property, as Friedman claims, and I does not satisfy the relevant instance of (*2*), as can be seen by substituting 'gred' for x.

We will now show that these defects cannot be repaired without change in the formal definitions. The following lemma reveals much of the content of Friedman's formalism:

Parasiticity Lemma

Let I be a predicate satisfying (*1*). Then

(*a*) $I \vee \neg C$ is minimally strong among the predicates satisfying (*1*) or any of its negation variants. Let us write J for $I \vee \neg C$.

(*b*) J satisfies the following negation variant of (*2*):

$$(K)[(x)\{Kx \rightarrow (\neg Jx \leftrightarrow Dx)\} \rightarrow (x)(Kx \leftrightarrow \neg Cx)]$$

Thus J is parasitic for D dependent on C, and if J is strictly weaker than I, I is *not* parasitic for D dependent on C.

The lemma of course remains valid under substitution of negations if done consistently throughout. The proof consists of straightforward logical manipulation and is left to the reader.

It is now easily verified that, for any predicates C and D, the predicate $D \vee \neg C$ satisfies (*1*); so by the lemma it represents information for D dependent on C. But any positive test for D depends on $D \vee \neg C$, for if the test implements a sufficient condition for D it implements a sufficient condition for $D \vee \neg C$ and hence is incompatible with the inapplicability of $D \vee \neg C$. Thus, in any putative proof that one of D and C is simpler than the other, any positive test for D is always excluded. Analogous consideration of $\neg D \vee C$ shows that any negative test for D is always excluded. The same holds for all tests for C, so according to Friedman's definitions all tests for any predicate are excluded and no predicate is simpler than any other.

KENNETH L. MANDERS
University of California, Berkeley

REFERENCES

FRIEDMAN, K. S. [1972]: 'Empirical Simplicity as Testability', *British Journal for the Philosophy of Science*, **23**, pp. 25–33.

FRIEDMAN, K. S. [1973]: 'Son of Grue: Simplicity vs. Entrenchment', *Noûs*, 7, pp. 366–78.

FRIEDMAN, K. S. [1975*a*]: to appear in *Mind*.

FRIEDMAN, K. S. [1975*b*]: private communication with the Author.

MANDERS, K. L. [1974]: 'Kenneth Friedman: Simplicity and Goodman's "New Riddle of Induction" ', unpublished.

ON WRITING POPPERIAN HISTORY

In his review article, 'Should Philosophers be allowed to Write History?',[1] L. Pearce Williams argues that the answer to his title question is 'No!' In particular he objects to the idea that Popper's 'doctrine of refutability' can be applied to the writing of history. He says:

> History is an inductive science. There are a finite number of facts that pertain to Faraday available to the historian. I feel secure in saying that it is highly unlikely that we will discover a new manuscript or letter that will radically change our knowledge of what Faraday did and wrote. That being the case, any attempt to understand Faraday and discover him, in Collingwood's sense, by getting inside his mind and following his ideas *must* take the whole body of historical facts on Faraday into consideration. The historian is not free to pick and choose those facts that suit his thesis, no matter how original or fascinating that thesis may be. Nor is it legitimate to make statements for which there is no evidence whatsoever just because one wishes to establish an idea that is essential to one's analysis. If one does, the result is not falsifiability and progress, but falsehood and bad scholarship. (p. 253)

There are several points that need to be made. It is not clear that the number of 'facts' concerning a given historical episode is fixed. In the light of a conjectured reconstruction, certain data which might seem irrelevant from one point of view, might be seen as relevant from the perspective of a new point of view. Without endorsing the procedure, one might consider psychoanalytic re-interpretations of history as accomplishing this feat. A theoretical perspective is capable of generating 'new facts' out of 'old data'.

Thus, even if we agree that the number of *facts* is fixed or limited, it does not follow that only *one* construal can be placed upon them. To the extent that any given set of facts can be construed in more than one way, the historian *must*, it would seem, construct a theory which coordinates the facts into a story. There is no reason why such theories cannot be falsifiable. Even if we assumed that the data from the past was *absolutely* fixed, one might select suitable subsets of the data, and using a conjectured theory, attempt to 'predict' the rest of the data. Thus, even if Pearce Williams is right in that no new facts about Faraday are likely to emerge, falsifiable history about Faraday is still possible.

Finally, Pearce Williams faults Berkson and Agassi with bad scholarship insofar as they select facts that support their theses and ignore other facts which do not. But, even if Pearce Williams's assessment of the merits of the attempts by Agassi and Berkson to write Popperian history is correct, this does not show that such histories should not be written, but only that the theories under review are bad theories.

MICHAEL BRADIE
Bowling Green State University

REFERENCE

WILLIAMS, L. P. [1975]: 'Should Philosophers be Allowed to write History?', *British Journal for the Philosophy of Science*, **26**, pp. 241–53.

[1] Williams [1975].

SYNCHRONISATION RULES AND TRANSITIVITY*

Adolph Grünbaum[1] has offered a proof of the intransitivity of non-standard signal synchronisms in inertial systems. This proof requires at least two signals, an indirect signal by which local, intermediate, and distant clocks are successively synchronised, and a signal directly connecting the local and distant clocks so that their resultant readings can be compared to check for transitivity. It further requires that the clock positions be non-collinear. Phillip Quinn[2] has extended Grünbaum's result to the collinear case in which the local clock is stationed between the others. Quinn contends, however, that the intransitivity of non-standard synchronisms depends crucially on the use of a direct signal, and describes a one-signal procedure which, he claims, produces transitivity.

Let A, B, C, respectively, be the local, intermediate, and distant clocks; l, m, n, respectively, the distances $\overline{AB}$, $\overline{BC}$, $\overline{AC}$; and ϵ_{AB}, ϵ_{BC}, ϵ_{AC}, the synchronisation rules. Quinn's basic idea is to require that a signal from A to C by which C's readings are compared with A's to check for synchrony be directed *via* B. If this is done, the reading of C upon arrival of the signal is $\epsilon_{AC}\,2\,(l+m)/c$ plus A's reading at departure. Suppose C to have been synchronised with B by a signal transmitted from B which departed from A when A read zero, and suppose that B was synchronised with A by the same signal. Then A's reading is the elapsed time since C was set, and that setting must have been $\epsilon_{AB}\,2\,l/c+\epsilon_{BC}\,2m/c$. Thus,

$$\epsilon_{AB}\,2\,l/c+\epsilon_{BC}\,2\,m/c = \epsilon_{AC}\,2\,(l+m)/c,$$

or

$$\epsilon_{AB}l+\epsilon_{BC}m = \epsilon_{AC}\,(l+m).$$

If B and C were synchronised by the same non-standard rule, this gives

$$\epsilon_{k}\,(l+m) = \epsilon_{AC}\,(l+m),$$

where $\epsilon_k = \epsilon_{AB} = \epsilon_{BC} \neq \frac{1}{2}$, so that $\epsilon_k = \epsilon_{AC}$.
Quinn therefore claims that the synchronism, although non-standard, is transitive, and infers that the transitivity properties of non-standard synchrony must be relativised to the choice of synchronisation procedure.

What transitivity is being relativised to, however, is not really the procedure used in setting the clocks, but the choice of path between A and C used to compare their readings to check for synchrony once C has been set. Of course, one could use the very signal by which B and C were set to compare C with A, but this is an incidental feature of Quinn's analysis. The most that can be concluded is that if the same non-standard rule is used to synchronise B with A and C with B, then whether or not C will be regarded as thereby synchronised with A according to that rule will depend on whether its readings are compared with A's *via* direct or circuitous signals.

In the case of standard synchrony, by contrast, this does not matter. The transitivity of standard synchrony is independent of the path of the signal by which previously set clocks are compared, since whether or not $l+m = n$, with $\epsilon_k = \frac{1}{2}$ we get $\epsilon_{AC} = \frac{1}{2}$ rather than $\epsilon_{AC} \neq \epsilon_k$ in Grünbaum's argument.[1] This is an excellent reason to adopt standard synchronism and, Quinn's argument notwithstanding, to regard $\epsilon = \frac{1}{2}$ as something more than a stipulation once synchrony has been required to be transitive. There is, after all, no good reason for disallowing the direct comparison of distant clocks.

---

* Research supported by a grant from the Excellence Fund of the University of North Carolina-Greensboro.

[1] Grünbaum [1969], pp. 9–11.

[2] Quinn [1974].

There is, however, some ambiguity in the view that non-standard synchrony is transitive under Quinn's restriction to indirect signals. Transitivity evidently requires that if B is synchronised with A and C with B according to a common rule for the setting of clocks, then C will thereby be synchronous with A if that same rule is used to compare its readings with A'*s*. Of course, C could (consistent with observation) be regarded as synchronous with A in any case, for there will always be some rule according to which they give identical readings. But if a different rule is required we do not regard the synchronism as transitive, but only as satisfying Grünbaum's *GRCS* condition.[2] If Quinn's indirect procedure for comparing C with A is regarded as a synchronisation rule, then to check for transitivity we would need to synchronise a fourth clock D with C by that same rule and then compare the resultant readings of D with those of A.

This is the view of I. W. Roxburgh,[3] with the qualification that Roxburgh wants to synchronise D with A by Quinn's rule and then, to check for transitivity, compare D with C. Roxburgh exacts this strategy from a passage in Reichenbach which I interpret differently.[4] Transitivity under Roxburgh's conditions would require that D and C turn out mutually synchronised only if their synchronisations with A had been required to be symmetric. And this requirement, he notes, would already impose standard synchrony. As I read Reichenbach, he is not, in the passage Roxburgh cites, attempting to provide a general definition of 'transitivity of synchrony', but an account of the advantages of the standard rule. Thus in comparing distant clocks which have been synchronised independently with a local clock to verify transitivity, Reichenbach is able to assume, and does so explicitly, that the synchronisms are symmetric. A general definition, applicable to non-standard cases, cannot carry this presupposition. If in fact Reichenbach intends the transitivity condition to mean that B and C will be mutually synchronised upon independent synchronisation with A, as Roxburgh supposes, then revision of his view would be called for. I therefore assume that if we follow Roxburgh in regarding Quinn's comparison procedure as a synchronisation rule, we must check it for transitivity by comparing D with A once it has been synchronised with C. That is, we evaluate the conditional:

$$CS_RA.\ DS_RC. \supset DS_RA,$$

where R is the rule that synchronisations proceed indirectly *via* B according to a fixed ϵ value.

The question is, how are the readings of D and A to be compared, once the antecedent has been satisfied? It is obvious that if they are compared *via* a signal transmitted directly to D from B upon arrival from A, the synchronism will turn out intransitive as Roxburgh contends. We would have

---

[1] Grünbaum [1969], equation 10′, p. 10.
[2] Grünbaum [1969], p. 8.
[3] Roxburgh [1975].
[4] Reichenbach [1958], p. 168.

$$\epsilon_{AC}\frac{2\,(l+n)}{c} = \epsilon\frac{2\,(l+m)}{c} + \epsilon\frac{2\,(m+n)}{c},$$

where $n = \overline{BD}$. This gives $(l+n)\,(\epsilon_{AC}-\epsilon) = \epsilon\, 2m$, with the result $\epsilon_{AC} \neq \epsilon$. This method of comparison may appear perfectly straightforward given the description we have been provided of R. Nevertheless, it does not seem to me to be the proper application to this situation of the procedure Quinn means to propose. The whole point of Quinn's procedure is to allow a *single signal*, appropriately transmitted and reflected, to accomplish both the setting of the successive clocks and the comparison of the resultant reading of the final clock in the sequence with that of the initial one. It is this that distinguishes Quinn's method from Grünbaum's, in which the final and initial clocks are compared *via* a signal path that bypasses the intermediate clock(s). Clearly, then, what Quinn means by a 'single signal' is not simply one wholly traceable to a single event at the initial clock (for in the present arrangement such a signal could be split at B with one part bypassing C and proceeding directly to D), but rather one whose path from initial to final clock includes the position of every intermediate clock and is otherwise direct. Quinn's synchronisation rule (if we are to follow Roxburgh in regarding Quinn's comparison procedure as a rule) is, therefore, that distant clocks be compared only over the signal path used in setting them. Under this restriction, the equation comparing D with A is:

$$\epsilon_{AD}\frac{2\,(l+2\,m+n)}{c} = \epsilon_{CD}\frac{2\,(m+n)}{c} + \epsilon_{AC}\frac{2\,(l+m)}{c}.$$

Assuming $\epsilon_{CD} = \epsilon_{AC} = \epsilon$, we get $\epsilon_{AD} = \epsilon$ even if $\epsilon \neq \frac{1}{2}$, in accordance with Quinn's position.

The important critical question, it seems to me, is how we are to assess Quinn's restricted comparison procedure. Allowing that it does issue in the transitivity of non-standard synchronisms, is it correct to regard it as a legitimate synchronisation rule? Once we recognise that Quinn's result, regardless of his mode of presentation, does not depend on the particular signal used to compare distant clocks, but only on the signal path, it is natural to regard Quinn's method not as a synchronisation rule at all, but only as an arbitrary restriction on the procedure for comparing clocks set independently according to some common choice of ϵ. The restriction is satisfiable but completely unmotivated. There is no reason to regard the transitivity of a synchronisation rule as a relation between the rule and a signal path rather than as a property of the rule itself. The natural view would be that an ϵ choice is intransitive if it requires a different ϵ choice for the identification of resultant readings on clocks compared directly, or, perhaps, *via* any path.

It is possible to claim that the 'natural view' is merely a convention, since in principle we are free to choose any signal path, including Quinn's. Quinn's accomplishment might then be represented as the revelation of a new level of conventionality—the conventionality of transitivity. It seems to me, however, that this sort of conventionality would have to be regarded as trivial even by those who insist that the choice of ϵ is non-trivially conventional. For the point of that view is to reveal the underdetermination of metrical relations by possible experience and the consequent freedom to make non-standard choices. Quinn's view that his non-standard synchronisms are transitive, by contrast, amounts to the imposition of a restriction under parallel conditions of freedom.

JARRETT LEPLIN
University of North Carolina at Greensboro

REFERENCES

Grünbaum, A. [1969]: 'Simultaneity by Slow Clock Transport in the Special Theory of Relativity', *Philosophy of Science*, 36, pp. 5–43.
Quinn, P. L. [1974]: 'The Transitivity of Non-Standard Synchronisms', *The British Journal for the Philosophy of Science*, **25**, pp. 78–82.
Roxburgh, I. W. [1975]: 'The Intransitivity of Non-Standard Synchronisms', *The British Journal for the Philosophy of Science*, **26**, pp. 47–48.
Reichenbach, H. [1958]: *The Philosophy of Space and Time*, Dover Translation.

Brit. J. Phil. Sci. 27 (1976), 403–414 *Printed in Great Britain*

Reviews

LEWIS, D. [1973]: *Counterfactuals*. Oxford: Basil Blackwell. £2.95. Pp. x+150.

This essay is a virtuoso performance in a certain way of doing philosophy. It represents an extension of recent work in the semantics of intentional logic, that which construes the necessity operators in terms of possible worlds. Lewis's concern is with the sentential connective that appears in what are traditionally called counterfactual conditionals, and this he introduces against the background of the usual possible-world semantics for modality. Strict conditionals are defined as material conditionals preceded by some necessity operator (*e.g.* $\Box(\phi \supset \psi)$, p. 14) and counterfactuals are exhibited as like yet unlike these. The technicalities may be intrinsically interesting, yet the philosophical interest lies in the explanatory power of such an approach and about this opinions will be divided. Doubtless possible-world semantics has a way of unifying a whole array of linguistic forms but inevitably many will find that in itself a success of doubtful value. Those who are not averse to such a technical approach may nevertheless query the pivot on which it turns, for possible worlds are not philosophically transparent entities.

To take a simple (and oversimplified) case, the counterfactual 'If kangaroos had no tails, they would topple over' is said to be true on the condition that kangaroos lacking tails *do* topple over in those possible worlds which are closest to our world. More generally, the analysis proceeds as follows. Our world is just one (the *actual* one) of the multitude of possible worlds and bears a varying degree of overall similarity to each of the rest. The non-actual worlds can be imagined to be arranged in a system of concentric spheres, with the actual world at the centre, in such a way that each sphere progressively further from the centre contains worlds which are overall less similar to our world than those in the sphere before. A counterfactual—a sentence of the form 'If it were the case that ϕ, then it would be the case that ψ', which Lewis writes as '$\phi \Box\!\!\rightarrow \psi$'—is true of (or at) our world on the condition that the closest antecedent worlds (in the smallest antecedent-permitting sphere) are all worlds in which the consequent holds. More fully yet, the analysis (p. 16) is this:

> $\phi \Box\!\!\rightarrow \psi$ is true at a world i (according to a system of spheres $\$$) if and only if either
>
> (1) no ϕ-world belongs to any sphere S in $\$_i$, or
>
> (2) some sphere S in $\$_i$ does contain at least one ϕ-world, and $\phi \supset \psi$ holds at every world in S.

Lewis's possible-world approach to counterfactuals is in a number of ways comparable to a theory developed by Stalnaker,[1] though it has the advantage over the latter in not assuming the existence of a *closest* possible world (pp. 77–83). The major rival to this approach is the theory (or family of theories) which Lewis calls 'Metalinguistic', developed by such philosophers as Chisholm, Goodman and Mackie: on that account the hypothetical $\phi \Box\!\!\rightarrow \psi$ is true (or in Mackie's version, valid) if suitable premises are available which together with ϕ allow the derivation of ψ. Lewis takes this rival seriously, even to the extent that he tries to eliminate the rivalry by deriving a version of it from his own analysis (pp. 65 ff.). This is one of the weak points of the book, not just in tending to obscure the differences between the approaches but also in its perverse construal of the notion of cotenability (p. 57). On the Metalinguistic account the 'suitable premisses' must be cotenable with ϕ, *i.e.* would not be ruled out by the truth of ϕ; on Lewis's definition however χ is cotenable with ϕ at i only if some ϕ-world is closer to i than any $\sim\chi$-world. It is of course an effective misconstrual. As for the differences these can be seen firstly from the fact that, from the point of view of a Metalinguistic theorist, Lewis would introduce irrelevant considerations in assessing a counterfactual. Lewis requires that some $\phi\cdot\psi$-worlds bear greater *overall* similarity to our world than any $\phi\cdot\sim\psi$-world if $\phi \Box\!\!\rightarrow \psi$ is true and this demands *all* features of these worlds are considered; in contrast, taking for example a counterfactual supported by laws, the rival account demands consideration only of those features of our world specifically relating to possible connections between the characteristics mentioned in ϕ and ψ. Secondly, the rival theories can assign different truth-values to at least some counterfactuals. Worlds which depart from our natural laws with the odd small miracle might be more similar to ours than some which stick rigidly to the laws but inevitably differ on many particular facts (pp. 75–7), and on such grounds Lewis's theory would allow miracles to disrupt precisely those ties between antecedent and consequent characteristics that the rival theory would take as establishing a counterfactual connection.

---

[1] Rescher (*ed.*) [1968].

Lewis takes a strong realist line on possible worlds. More philosophers might find a possible-world analysis of counterfactuals acceptable as an interim account, with a reductive analysis to follow in terms perhaps of sets of sentences or behaviourally (as Mackie suggests) of supposing. Lewis is uncompromising: more than just a primitive notion in his analysis, possible worlds are irreducible, 'respectable entities in their own right'. Metaphysically suspect as they are either for reasons of obscurity or parsimony, Lewis makes a brave defence of his realism claiming that it is forgivable in being only quantitatively and not qualitatively unparsimonious. 'You believe in our actual world already. I ask you to believe in more things of that kind, not in things of some new kind' (p. 87). One problem facing such a view is the question of the inhabitants of such other worlds—are they identical with those of our world? Lewis suggests that individuals (if not universals) inhabit only one world, but have *counterparts* in others (pp. 39 ff.). This leads in fact to a complication in the official account of counterfactuals (p. 41) which appears to miss official recognition, for only one other reference is made to it in the book.

The second fundamental notion in the account is that of comparative similarity, about which Lewis is interesting and persuasive. The notion may well be accused of vagueness, yet it is vague 'in a well-understood way' (p. 91). Moreover, counterfactuals are also undeniably vague in their truth conditions. 'Overall similarity consists of innumerable similarities and differences in innumerable respects of comparison, balanced against each other according to the relative importances we attach to those respects of comparison' (p. 91). One arrangement of spheres around a world *i* represents but one overall assessment of similarity, in terms of which $\phi \Box\!\!\rightarrow \psi$ can be judged true; another permissible arrangement might make $\phi \Box\!\!\rightarrow \psi$ at *i* false—but this is as it should be. Not all arrangements are permissible indeed (p. 93) but no one is mandatory. Such is Lewis's treatment of the 'semantic indeterminacy' or 'pragmatic ambiguity' of counterfactuals as it was variously recognised by (among others) Chisholm, Rescher and Stalnaker.

Two consequences of Lewis's analysis deserve specific comment. In the case where ϕ is impossible, $\phi \Box\!\!\rightarrow \psi$ is vacuously true for any ψ (p. 24). This is I think an unfortunate consequence because an account of counterfactuals should discriminate (as we do) between what is and what is not counterfactually implied by impossible antecedents. 'If the contents of this jar were a body of ideal gas, it would obey Boyle's Law' is one we might favour even though it could not be, just as there can be no frictionless planes nor perfectly elastic bodies. The other salient consequence is that $\phi \Box\!\!\rightarrow \psi$ is true if ϕ and ψ are both true (pp. 26–9), a point with which the Metalinguistic rival will take issue. In a temporary spirit of conciliation Lewis offers an alternative account to the official one whereby the system of spheres around *i* is not centred on *i* itself, thereby dropping the assumption that no other world is as similar to *i* as *i* itself (p. 29). The official account nevertheless lives on.

There is much more to the book than the skeleton sketched above to satisfy the largest appetite for definitions, transformations and regimentations. As well as the 'would' operator $\Box\!\!\rightarrow$ chapter 1 introduces the 'could' operator $\Diamond\!\!\rightarrow$ defined in terms of the former, and with their help defines also the usual modal operators $\Box$ and $\Diamond$ (which Lewis calls 'outer modalities'). Chapter 2 on 'Reformulations' is devoted to a number of alternative ways of presenting the official analysis. Chapter 5 on 'Analogies' is an intriguing description of some notions that bear close formal analogies to counterfactuals: possible worlds might be nested in terms for example of comparative goodness so that $\Box\!\!\rightarrow$ could be read as the conditional obligation operator of deontic logic, and an attempt is made to use the formal apparatus to give an account of certain kinds of definite descriptions. The work is rounded off of course with a chapter of completeness proofs.

BRIAN CARR
University of Exeter

REFERENCE

RESCHER, N. (*ed.*): [1968]: *Studies in Logical Theory*. American Philosophical Quarterly Monograph No. 2.

AGASSI, J. [1975]: *Science in Flux*, *Boston Studies in the Philosophy of Science*, **28**, edited by R. S. Cohen and M. W. Wartofsky. Dordrecht: Reidel. Pp. xxvi+553.

The present volume is a collection of twenty-odd essays on the metaphysical, factual and social aspects of science. Most are reprinted. There is so much redundancy and scatter in the volume that I fear many readers will not discover the challenging ideas which it contains. A more stringent editorial policy would have benefited the philosophical community and conserved trees. But let me begin with the interesting philosophy and complain about the format later.

In an essay published in 1963[1] Popper added a famous 'third requirement' to his characterisation of the growth of knowledge. In order that there be science as we know it, not only must scientists make bold explanatory conjectures and test them severely, there must also be corroborations. Our theories must not always be refuted immediately. Thus, as Agassi implies, the book might better have been called *Conjectures, Corroborations and Refutations*.

In this volume Agassi develops a more radical and austere position—one which *may* be present in Popper's earlier works although Agassi is not quite sure. As he rather plaintively remarks after wrestling with this exegetical problem, 'I liked Popper better before I understood him'. (p. 386.) Agassi's position, as I understand it, is this. We gain *theoretical* knowledge about the world through experiments which refute our conjectures. Contrary to Popper's third requirement, no successful predictions are necessary for theoretical science. When Agassi uses the Popperian slogan, 'We learn from our mistakes', he means it in a very strong literal sense.

For example, from Agassi's position it would follow that Stahl had a good deal of theoretical knowledge about acids because he knew that not all acids were organic in origin (there were mineral acids, too), not all acids contained phlogiston (it was only after sulphur underwent combustion that it became acidic), and not all acids would attack metals. Likewise today scientists have theoretical knowledge of schizophrenia in as much as they know a lot of things which don't cause it.

I am not comfortable calling such information from refuting experiments *theoretical* knowledge, yet I was surprised at how good a case Agassi was able to make for his theory of *scientia negativa*. The title essay, 'Science in Flux', deals with one *prima facie* difficulty for this position. If Agassi (following the early Popper) is right in stressing the overwhelming importance of refutation for the growth of science—as he puts its, 'Criticism belongs not to the *hors d'oeuvre*, but to the main dish' (p. 33)—why is it that actual science, including physics, is so remarkably stable? Why aren't there revolutions all the time? Agassi offers three quite different reasons for the theoretical sluggishness in science: First of all, not many of us are imaginative enough to invent really bold new conjectures and what creativity we do have is often stultified by the educational system and institutionalised scholasticism. Secondly, science is stabilised by metaphysical frameworks which are difficult to criticise and which therefore tend to persist. Thirdly, much of what is called science is really technology or applied science and in these domains what Agassi calls 'positive evidence' *is* important. It is legitimate for a technologist to explore at length the domain of practical applicability of an already refuted theory.

At this point the reader is probably starting to relax. 'Oh', you are saying, 'confirmations count after all—it's just that Agassi relegates them to what he calls applied science'. But this would be to misunderstand the deep scepticism of Agassi's position. For Agassi, bits of positive evidence (what others might call confirmations or corroborations) have *no* probative force, as far as we can tell. Their function is not epistemological, but social and moral. We need positive evidence, not in order to act wisely, but in order to act *responsibly*.

---

[1] Popper [1963].

Agassi's theory about the role of positive evidence is developed in a series of essays on philosophy of technology (see especially, 'Positive Evidence as a Social Institution'). The basic argument, as I understand it, is this: No prediction can ever receive genuine inductive support. However, if we adopt some comprehensive theoretical system such as Newtonian mechanics or modern chemistry, then we can have induction by elimination and Bayesian confirmation within that framework. So Agassi does not deny the possibility of local or relative inductive support. But he does deny that the theoretical system itself can be justified. Nevertheless, because people must act and plan and coordinate efforts, societies adopt by convention certain methods for evaluating and applying scientific frameworks. These social conventions can be criticised and improved but they cannot provide any sort of epistemological foundations for decisions.

Thus, it is a convention in our society to collect positive evidence concerning the efficacy and safety of drugs before prescribing them. We cannot rationally claim that one is less likely to endanger the patient receiving the drug if one tests it first. But we can say that it is irresponsible not to subject it to the best tests one can think of. Trying to eliminate the most obvious errors in our theories before acting on them cannot relieve our post-Human intellectual anxieties, but at least it gives us a clear conscience.

Agassi accepts Hume's philosophical analysis of induction but replaces Hume's descriptive psychology with a sociological theory of inductive reasoning which restores to it a certain amount of intellectual respectability. On Agassi's account conventional inductive rules are not psychological inevitabilities but proposals which can be subjected to rational public debate. His programme for understanding induction would involve us in historical and sociological enquiry and not in a hopeless quest for *the* inductive logic. For example, we might study in detail the effects of the methodological rules of the early Royal Society on research. Or we could compare the quality control standards of various societies. Was it a lucky accident, asks Agassi, that thalidomide was not used in America, or was it the result of an articulated policy? How do standards affect the rate of innovation? These are the sorts of questions which philosophers should be investigating.

What emerges is a picture of science in which there are no fixed elements. Even the canons of rationality are open to criticism. But if there are no epistemological foundations, what provides the basis for criticism? Agassi makes it very clear that even that form of criticism which seems most straightforward, *viz.* empirical testing, should be viewed as a 'bootstrap operation'. What is characteristic of a bootstrap operation is that we use one conjectural statement which is itself either false or dubious to criticise and improve a second conjecture. He sketches several examples of this process. 'The perfect gas law [was used] to develop the kinetic theory of gases without which Van der Waals could not possibly have developed his replacement of the perfect gas law' (p. 170). And Newton used Kepler's laws to calculate the masses of the planets which he then plugged into his theory in order to show that the solar system does not obey Kepler's laws exactly. Agassi does not analyse the logic of any of these situations in detail. In fact his comment after listing several such examples is very typical of the philosophical style of the entire book: 'If these kinds of practice do not feel like a bootstrap operation, if all this does not give the logically minded reader a slight shudder, then that's the end to the problem: he can just as well ignore it or delight in his solution to it as the case may be' (p. 171).

Agassi seems very concerned not to bully or bore his reader with detailed analyses or arguments. As a result there are allusions to many exceedingly interesting philosophical possibilities and a great number of striking images and aphorisms, such as 'The Naked Truth does not carry a handbag containing identification, corroboration, or letters of recommendation' (p. 421), but very few of the theses are carefully developed.

I, for one, would find this way of doing philosophy an acceptable alternative if the literary style of the writing were pleasing. But it often is not. There is a considerable overlap between the essays—to give an extreme example, pages 333–4 repeat pages 316–18 *verbatim*. Single essays are often rambling (see especially 'Unity and Diversity in Science' and 'Modified Conventionalism') and even individual sentences are often confusingly ambiguous. For example, the essay entitled 'When Should We Ignore Evidence In Favour of a Hypothesis?' is *not* about ignoring confirmations, as this reader at least first supposed, but about keeping hypotheses in the face of anomalies.

Agassi writes, 'I consider Boyle's ruling, in which he proclaimed that he would ignore the publications of chemists who would not describe their experiments in a sufficiently clear manner to render them repeatable, very wise, indeed' (p. 383). Agassi's own writing is obviously clear enough for me to have attributed a position to him. But it is needlessly obscure and unstructured. A little more effort on the part of the author and editors would have made life easier for both reviewers and prospective readers.

NORETTA KOERTGE
Indiana University

REFERENCE

POPPER, K. R. [1963]: 'Truth, Rationality, and the Growth of Scientific Knowledge', in *Conjectures and Refutations*, pp. 215–50. London: Routledge and Kegan Paul.

VLASTOS, G. [1975]: *Plato's Universe*. Clarendon Press: Oxford University Press. Pp. xiii+130. £3.75.

This is an admirably clear, brief account of the Presocratic background to Plato's cosmology and of two of the principal items in that cosmology, his doctrines on celestial motion and on the structure of matter. Professor Vlastos focuses attention on the Presocratic notion of a world-order. *Physis*—nature—he writes 'is the key term in the transition from the world of Homer, Hesiod, Archilochus, Sappho, Aeschylus, Sophocles, Aristophanes, Herodotus, and the orators' (a list of names that indicates the spread of material he draws on) . . .' to the world of the *physiologoi* and of a few tough, hard-bitten, intellectuals like Thucydides and the Hippocratics—the world which was cosmos . . . The beauty of *physis* is that it shows what it was in the traditional conception of the world that gave the *physiologoi* the building materials for their new construction. The cosmos they had to invent.' Plato he sees as taking over this idea of cosmos from the natural philosophers while disagreeing fundamentally with them on the questions of the priority of soul to matter and of the role of god. His world-system, like that of the earlier cosmologists, is one 'safe from supernatural interference': but now the world-order itself depends upon a creator-god. 'If you cannot expunge the supernatural, you can rationalise it, turning it paradoxically into the very source of the natural order, restricting its operation to a single primordial creative act which insures that the physical world would be not chaos but cosmos forever after.' Plato thus 'in his own perversely original way, sustained the faith that our world is cosmos'. Indeed, according to Professor Vlastos, 'he gave rational men a pious faith to live by in two millennia all through which science was more prophecy than reality'. But there was more to Plato's contribution even than that. Professor Vlastos underlines the importance of the conception that the apparent courses of the heavenly bodies can be explained in terms of combinations of circular movements. The resulting theory 'would have been of special interest to the working astronomer', as he puts it, 'because it generated a multitude of inferences concerning the sun's expected behaviour—inferences which could be used to institute new observations that might confirm or disconfirm the theory'. Here then he sees Plato as preparing the way for the triumph of Eudoxus' theory of concentric spheres. Finally in his discussion of Plato's theory of matter he emphasises that this—like its rivals—was 'safe from observational refutation': to refer to empirical data here as the criteria between competing theories is to misconstrue the nature of the controversy; Plato's aim was, rather, a theory that was both credible and, more important, aesthetically satisfying.

Sketched thus broadly, Professor Vlastos's overall thesis concerning Plato and his response to the Presocratics is one we should, in the main, wholeheartedly endorse, and it has the merits not just of lucidity and succinctness, but of drawing attention to neglected aspects of these topics. Nevertheless some reservations must be expressed. Two relate to his discussion of the Presocratic background. Professor Vlastos writes, like Aristotle, of the *physiologoi* or natural philosophers in general, a term that covers thinkers as divergent in their views as the Milesians, Heraclitus, Empedocles, Anaxagoras, Diogenes of Apollonia and the fifth-century atomists, Leucippus and Democritus. He is, of course, well aware of the disagreements among these philosophers not just on the problem of the material constitution of the universe, but also on such issues as whether the world is created, whether there is one cosmos or many, on the nature of god and on the role (if any) of a creative intelligent agency in the world. But in practice he tends to play down these differences. Yet when Plato castigates the atheists who believed that everything comes to be 'by nature' and 'by chance' as opposed to 'by art' or by design, in *Laws* X—a key text to which Professor Vlastos rightly gives prominence—the question of who precisely is condemned is problematic. Professor Vlastos sees Plato's target as the *physiologoi* in general, for he had a quarrel even with those who believed in a divine intelligence in the universe (Heraclitus, Anaxagoras, Diogenes of Apollonia) in that they saw order as immanent in nature, not the result of a transcendent force. Yet in the text of the *Laws* at least intelligence (*nous*) as such is placed on the side of 'art' and 'god', and there is no hint of a contrast between immanent and transcendent intelligences. Professor Vlastos represents Plato as reacting to a whole tradition (the 'declared enemy and would-be persecutor of the *physiologoi*') where it could be argued that *Laws* X at least is directed more narrowly at one branch of that tradition.

A second and perhaps more substantial doubt concerns the way the contrast between the rational and the irrational is deployed. Here Professor Vlastos concentrates on the question of appeals to the supernatural in accounts of natural phenomena—a mode of explanation that can be illustrated in Herodotus, for instance, as well as in Homer and Hesiod. In contrast to Herodotus, the natural philosophers 'make the world a cosmos by keeping what was already there in the form of *physis* and cleaning out everything else.' This move, as he goes on to note, is implicit in the philosophers, though one might question the observation that 'to attack the supernatural head-on would turn them into outlaws' when such explicit attacks do occur in the medical writers. But more importantly, it is not as if, by excluding references to the supernatural in their accounts of earthquakes, thunder and lightning and so on, the philosophers thereby created a world in which the irrational simply had no purchase whatever. This is not just a matter of the presence, among the *physiologoi*, of a man such as Empedocles, who was as much concerned with the salvation of the *daimon* as with the constitution of matter: it is that there is more to the Presocratics' speculation on such topics as god and the soul than the mere rationalisation or naturalisation of earlier ideas. If we take Democritus himself, the star example of a natural philosopher, it is clear that he explained the gods, dreams and so on in terms of atoms and their 'images' (*eidola*). Yet within this materialist framework he appears to have accommodated some traditional beliefs. Thus our sources suggest that he admitted the possibility of 'images' revealing the future and that he desired 'propitious' 'images' for himself, and although some of our evidence in this area is questionable, not all of it can be dismissed. The importance of the philosophers' achievement is undeniable: but one might argue that it was more limited in scope, and less clear-cut, than some of Professor Vlastos's remarks may suggest.

In interpreting Plato's astronomy, Professor Vlastos lays great stress on the originality of the idea that the composition of postulated regular circular motions may account for irregular phenomenal motions and on the continuity of Plato's work with that of Eudoxus, who was, as no one doubts, responsible for the first detailed mathematical model for celestial motion. In Professor Vlastos's view Plato had already provided a solution for the sun, though it was left to Eudoxus to do so for the planets and moon. The problem here is simply one of a lack of information first concerning the details of pre-Platonic astronomy (especially the Pythagorean system later associated with the name of Philolaus) and then about Plato's own relations with his contemporaries. So far as our extant evidence, goes, Plato is certainly responsible for the first clear statement of two important ideas, (1) the contrast between the two main movements (those of the circles of the Same and of the Other), westerly round the poles of the celestial equator and easterly along the zodiac, and (2) the connection between the movements of Venus and Mercury on the one hand and of the sun on the other. What is not at all clear is how far Plato was drawing on the ideas of his predecessors or contemporaries. Now the importance of Plato's statements, in the *Republic*, concerning the proper role of astronomy in the education of the guardians is undeniable, even though the interpretation of those statements remains highly controversial. Again the evidence for his setting his pupils the task of resolving irregular apparent motions into regular ones will be rejected only by the fairly hardened sceptic—even though that evidence is indirect. But when it comes to his contribution towards solving specific problems, the element of doubt increases. Plato's growing confidence in handling the topic of celestial motions is apparent from a comparison of the *Phaedo*, the *Republic*, the *Timaeus* and the *Laws*, but how much that confidence is due to the success of his associates' work—notably that of Eudoxus himself—is impossible to determine. So far as the *Timaeus* goes, however, the elliptical and in parts very obscure account of the relations between sun moon and planets (*e.g.* 36d), his drawing back from a discussion of the outer planets at 38de, and his suggestion, at 40cd, that to deal with such matters as the 'reversals and progressions' of the heavenly bodies without recourse to visible models is wasted labour, all read more like the remarks of someone who is excerpting those parts of current astronomical theory that he needs for his own cosmological purposes, rather than of someone actively engaged himself in tackling astronomical problems. This is, of course, a subjective impression: but the evidence to clinch the matter one way or the other does not exist—which in turn means that the confidence with which we can speak of Plato's original contributions to astronomical theory—as opposed to what we might call the philosophy of astronomy—has necessarily to be muted.

On the question of the structure of matter, Professor Vlastos again concentrates on one of the basic issues, the role of observational data, and argues forcefully that Plato's doctrine, like its rivals, is immune to empirical refutation: whilst all early Greek physical theories attempted *some* account of what were accepted as the main 'facts', none was falsifiable by reference to such data, and the main grounds on which one theory was preferred to another were often aesthetic. The distinction that Professor Vlastos here insists on, between *accommodating* what passed as the facts and actively seeking to *confront* new observational data, is the crucial one for our understanding of Greek physical theory. One would, however, have liked him to extend his inquiry to include an analysis of the role, in the fifth and fourth centuries, of arguments such as those concerning divisibility recorded by Aristotle. Dealing with the texts in Aristotle in which Plato's theory is criticised for failing to take the *phainomena* into account, Professor Vlastos shows very clearly how ineffective those criticisms were: from another point of view, however, those texts provide evidence of a sort for *attempts* to appeal to empirical factors.

Within a brief compass, Professor Vlastos has covered a great deal of ground, tackling many issues in considerable detail both in the text and in a series of appendices devoted to specific points. He has brought the question of the relationship between Plato and the Presocratics—and of the continuities and discontinuities in fifth- and fourth-century cosmology, astronomy and physics—into sharp focus, and provided student and specialist alike with an informative and thought-provoking guide to a set of problems that mark a turning-point in the development of early Greek science.

G. E. R. LLOYD
King's College, Cambridge

SOBER, ELLIOTT [1975]: *Simplicity*. Oxford: Clarendon Press. (Clarendon Library of Logic and Philosophy.) £6.50. Pp. 189.

That the simplicity of a scientific theory is grounds for 'accepting' the theory is a thesis held by a very wide spectrum of scientists and philosophers of science. But what is it for one theory to be simpler than another, and why is the simplicity of a theory grounds for accepting it? These resultant questions are highly controversial ones, and it is with the answer to them, and in particular to the first question, that Dr Sober's book is concerned. The main thesis of the book is that the simplicity of a theory is a matter of its informativeness, which in turn is a matter of how much extra information has to be obtained to enable the theory to answer questions of interest to the scientist. The less extra information you need to obtain before the theory will answer your questions, the more informative is the theory. Thus 'all swans are white' is a simpler theory than 'all swans are white, except Australasian swans which are black', because the extra information 'this is a swan' would enable the first theory to answer the question 'what colour is it?', whereas we need the further information of where the swan lives before the second theory is able to answer that question.

The view that the simplicity of a scientific theory is basically a matter of its informativeness is of course to be found in chapter 7 of *The Logic of Scientific Discovery*. However Sober interprets the informativeness of a theory not in terms of the ease of falsifying it, but in the way described above. It is not at all obvious to me that Sober's way of understanding 'informativeness', rather than Popper's way or some other way, is the obvious way to understand that notion. Are not both the above theories in a sense equally informative, for both tell us what is the colour of each swan in the world? However, given his understanding of 'informativeness', Sober develops it in a careful and rigorous way, and shows how it can account for certain judgments of simplicity which we make, which it might seem initially that a Popperian thesis could not adequately cope with. Thus, what makes $y = 3x+6$ a simpler theory than $y = 5x^4+4x^3-75x+168$? Both theories need to have added to them merely the value of x in order to answer the question 'what is the value of y?' So why is the first one simpler? Popper's answer in terms of the 'generality' of the former theory seems to bring in considerations other than informativeness (in his sense). The merit of Sober's account is that his answer to this question remains an answer in terms of informativeness (in his sense). This answer is that although both theories need the same amount of extra information to answer the question 'what is the value of y?', they do not require the same amount of extra information to answer the question 'what is the value of dy/dx'. The first theory needs no extra information to tell you this, whereas the second theory needs to have added to it the value of x (or the value of y). A similar answer accounts for the greater simplicity of a circle hypothesis over a rival ellipse hypothesis. Of course Sober's account does have the consequence that the simplicity of a scientific theory is question-relative; and, since which questions can be asked is a function of which predicates a scientific community has available to it, simplicity is relative to vocabulary. Sober accepts this consequence, and admits that 'all emeralds are green' is simpler than 'all emeralds are grue' only relative to our vocabulary, and so to the kinds of questions which we ask, and not relative to the vocabulary of a grue-speaking community. The former theory needs less extra information to enable it to answer the question 'is it green?' than does the latter theory, but the latter theory is better at answering the question 'is it grue?'.

The opening chapter of the book expounds the main thesis, and the second chapter draws out some of its consequences, such as those which I have just sketched. It makes a plausible case for the claim that Sober's thesis captures many of our intuitions about simplicity. These two chapters are illuminating and challenging—even if not completely convincing. (Surely a man would prefer $y = 3x+6$ to $y = 5x^4+4x^2-75x+168$, even if we had never heard of a derivative?). In chapters 3 and 4 Sober then applies this thesis to two areas of 'science', and shows its consequences in those areas. This clearly needs to be done in a book about simplicity; the theory needs to account for the basic judgments of simplicity which scientists make. But here something very odd happens. Instead of attempting to show that his theory accounts for the judgments made in some paradigm area of science, in some branch of physics, or chemistry, astronomy or microbiology, Sober attempts to show that it works in transformational phonology, and in the psychology of perception. There are certainly not central areas of science; in the case of each it is open to question whether we have a science at all, or at any rate a science of the normal type. In neither case does the practitioner of the 'science' attempt to produce generalisations of the exceptionless character of laws of nature, and in both cases he attempts to produce theories which themselves account for the judgments made by conscious human beings (*e.g.*, by language users about which utterances are grammatical). So even if Sober's theory does work in these areas (and I think that it does), his conclusions do not for that reason necessarily apply to central areas of science.

Where this book however is least satisfactory is in what it says about the place of simplicity in science. Although it gives a plausible account of what the simplicity of a scientific theory consists in, it does not give a plausible account of how the simplicity of a scientific theory competes with other criteria for its acceptability, and why simplicity is a criterion of acceptability. Indeed the book gives the impression that Sober has not thought very deeply about these matters. To start with, he talks a lot for many pages about 'accepting' or 'choosing' a scientific theory without saying what it is to 'accept' or 'choose' a scientific theory. The reader wonders whether to 'accept' is to believe to be true, or to use as a basis for action, or what. Sober does eventually remove puzzlement on this score, by remarking towards the end of the book (p. 154) 'we choose hypotheses in part for their simplicity, where "choose" means "reasonably regard as true"'. However, the unwary should not conclude from this that Sober regards the simplicity of a scientific theory as evidence of its truth. Indeed, with a very slight qualification, Sober explicitly rejects this view, a view which he calls 'probabilism'. His view is that there are two criteria for the acceptability of a scientific theory—simplicity and 'support'. We seek a theory which has 'support' because such a theory is 'relatively safe', and one which is simple because simple theories are informative. We seek simple theories because we want information; we seek 'supported' theories because we want that information to be true.

Unfortunately Sober gives no account of the other criterion of 'acceptance'—

'support'. He seems to feel that this is another topic. His general view is that (p. 168) 'support and simplicity are irreducibly distinct goals in hypothesis choice'. But he does need to make clear just what this other criterion is, in order to show what are the boundaries and role of his own criterion of simplicity. He begins by saying that a theory is 'supported' to the extent to which evidence favours it. But if that is what support is, it is highly plausible to suppose that the simplicity of a hypothesis H affects the value of its degree of support on evidence E. 'All swans are white' is better supported on the evidence that a thousand swans have all been found to be white, than is 'all swans are white, except Australasian swans which are black', and that is just because it is simpler. But a lot of Sober's discussion suggests that he means something else by 'support'. Thus, on page 74, considering the choice between rival equations, he interprets 'support' as 'goodness of fit'. This is a much narrower concept of support and would allow him to ask the question which he asks on page 166, and which he seems to think is an interesting question: 'take two theories that seem to be equally supported by the evidence. What reason is there to suppose that the simpler of the two is more likely to be true?'. If both theories are equally 'supported' in that the evidence favours each equally, then it is tautologous that each is equally likely to be true, and there would be no point in Sober asking this question.

So because Sober has not thought much about 'support', he has not given a satisfactory account of whether and to what extent simplicity is evidence of truth, and so of why the scientist seeks for simple theories. He may indeed seek for simple theories because such theories are more informative in Sober's sense; but it is highly plausible to suppose that as in the example in my last paragraph, he seeks simpler theories, also because a simpler theory is more likely to be true—more likely, that is, than some other theory of equal scope. (By theories of equal scope I mean theories which give us the same kind and amount of information about the same objects, *e.g.*, theories which tell us the colour of every swan in the world. Two theories of equal scope are equally informative on a different understanding of informativeness from Sober's.)

RICHARD SWINBURNE
University of Keele